

GENERALIZED NETWORK TOMOGRAPHY *

GUGAN THOPPE†

Abstract. Generalized network tomography (GNT) deals with estimation of link performance parameters for networks with arbitrary topologies using only end-to-end path measurements of pure unicast probe packets. In this paper, by taking advantage of the properties of generalized hyperexponential distributions and polynomial systems, a novel algorithm to infer the complete link metric distributions under the framework of GNT is developed. The significant advantages of this algorithm are that it does not require: i) the path measurements to be synchronous and ii) any prior knowledge of the link metric distributions. Moreover, if the path-link matrix of the network has the property that every pair of its columns are linearly independent, then it is shown that the algorithm can uniquely identify the link metric distributions up to any desired accuracy. Matlab based simulations have been included to illustrate the potential of the proposed scheme.

Key words. generalized network tomography, generalized hyperexponential distributions, unicast measurements, moment estimation, polynomial systems

AMS subject classifications. Primary, 47A50; Secondary, 47A52, 62J99, 65F30, 65C60, 68M10, 68M20

1. Introduction. The present age Internet is a massive, heterogeneous network of networks with a decentralized control. Despite this, accurate, timely and localized information about its connectivity, bandwidth and performance measures such as average delay experienced by traffic, packet loss rates across links, etc. is extremely vital for its efficient management. Brute force techniques, such as gathering the requisite information directly, impose an impractical overhead and hence are generally avoided. This necessitated the advent of network tomography—the science of inferring spatially localized network behaviour using only end-to-end aggregate metrics.

Recent advances in network tomography can be classified into two broad strands: i) traffic demand tomography—determination of source-destination traffic volumes via measurements of link volumes and ii) network delay tomography—link parameter estimation based on end-to-end path level measurements. For the first strand, see [27, 15, 29]. Under the second strand, the major problems studied include estimation of bottleneck link bandwidths, e.g. [18, 9], link loss rates, e.g. [3], link delays, e.g. [7, 22, 23, 26, 5, 8], etc. Apart from these, there is also work on estimation of the topology of the network via path measurements. For excellent tutorials and surveys on the state of the art, see [1, 7, 4, 14]. For sake of definiteness, we consider here the problem of network delay tomography. The proposed solution is, however, also applicable to traffic demand tomography.

Given a binary matrix A , usually called the path-link matrix, the central problem in network delay tomography, in abstract terms, is to accurately estimate the statistics of the vector X from the measurement model $Y = AX$. Based on this, existing work can be categorized into deterministic and stochastic approaches. Deterministic approaches, e.g. [11, 6, 10], treat X as a fixed but unknown vector and use linear algebraic techniques to solve for X . Clearly, when no prior knowledge is available, X can be uniquely recovered only when A is invertible, a condition often violated in practice. Stochastic approaches, e.g. [5, 23, 26, 29], on the other hand, assume X to

*A summary of this work without proofs has appeared in the proceedings of the 50th Annual Allerton Conference on Communication, Control and Computing (see [25]).

†School of Technology & Computer Science, Tata Institute of Fundamental Research, 1st Homi Bhabha Road, Mumbai, INDIA (gugan@tcs.tifr.res.in).

be a non-negative random vector of mutually independent components and employ parametric/non-parametric estimation techniques to infer the statistical properties of X using samples of Y . In this paper, we build a stochastic network tomography scheme and establish sufficient conditions on A for accurate identification of the distribution of X .

Stochastic network tomography approaches, in general, model the distribution of each component of X using either a discrete distribution, e.g. [26, 29], or a finite mixture model, e.g. [5, 23]. They construct an optimization problem based on the characteristic function, e.g. [5], or a suitably chosen likelihood function, e.g. [23, 26, 29], of Y . Algorithms such as expectation-maximization, e.g. [23, 26, 29], generalized method of moments, e.g. [5], etc., which mainly exploit the correlations in the components of Y , are then employed to determine the optimal statistical estimates of X . In practice, however, these algorithms suffer two main limitations. Firstly, note that these algorithms utilize directly the samples of the vector Y . Thus, to implement them, one would crucially require i) end-to-end data generated using multicast probe packets, real or emulated, and ii) the network to be a tree rooted at a single sender with destinations at leaves. Divergence in either of the above requirements, which is often the case, thus results in performance degradation. Secondly, the optimization problems considered tend to have multiple local optima. Thus, without prior knowledge, the quality of the estimate is difficult to ascertain.

In this paper, we consider the problem of generalized network tomography (GNT) wherein, the objective is to estimate the link performance parameters for networks with arbitrary topologies using only end-to-end measurements of pure unicast probe packets. Mathematically, given a binary matrix A , we propose a novel method, henceforth called the distribution tomography (DT) scheme, to accurately estimate the distribution of X , a vector of independent non-negative random variables, using only IID samples of the components of the random vector $Y = AX$. In fact, our scheme does not even require prior knowledge of the distribution of X . We thus overcome the limitations of the previous approaches.

We rely on the fact that the class of *generalized hyperexponential* (GH) distributions is dense in the set of non-negative distributions (see [2]). Using this, the idea is to approximate the distribution of each component of X using linear combinations of known exponential bases and estimate the unknown weights. These weights are obtained by solving a set of polynomial systems based on the moment generating function of the components of Y . For unique identifiability, it is only required that every pair of columns of the matrix A be linearly independent, a property that holds true for the path-link matrix of all multicast tree networks and more.

The rest of the paper is organized as follows. In the next section, we develop the notation and formally describe the problem. Section 3 recaps the theory of approximating non-negative distributions using linear combinations of exponentials. In Sections 4 and 5, we develop our proposed method and demonstrate its universal applicability. We give numerical examples in Section 6 and end with a short discussion in Section 7.

We highlight at the outset that the aim of this paper is to establish the theoretical justification for the proposed scheme. The numerical examples presented are only for illustrative purposes.

2. Model and Problem Description. Any cumulative distribution function (CDF) that we work with is always assumed to be continuous with support $(0, \infty)$. The moment generating function (MGF) of the random variable X will be $M_X(t) =$

$\mathbb{E}(\exp(-tX))$. For $n \in \mathbb{Z}_{++}$, we use $[n]$ and S_n to represent respectively the set $\{1, \dots, n\}$ and its permutation group. We use $\binom{n}{k_1, k_2, \dots, k_d}$ to represent $\frac{n!}{k_1! \dots k_d!}$. Similarly, $\binom{n}{k}$ stands for $\frac{n!}{k!(n-k)!}$. We use the notation \mathbb{R}, \mathbb{R}_+ and \mathbb{R}_{++} to denote respectively the set of real numbers, non-negative real numbers and strictly positive real numbers. In the same spirit, for integers, we use \mathbb{Z}, \mathbb{Z}_+ and \mathbb{Z}_{++} . All vectors are column vectors and their lengths refer to the usual Euclidean norm. For $\delta > 0$, $B(\mathbf{v}; \delta)$ represents the open δ -ball around the vector \mathbf{v} . To denote the derivative of the map f with respect to \mathbf{x} , we use $\dot{f}(\mathbf{x})$. Lastly, all empty sums and empty products equal 0 and 1 respectively.

Let X_1, \dots, X_N denote the independent non-negative random variables whose distribution we wish to estimate. We assume that each X_j has a GH distribution of the form

$$(2.1) \quad F_j(u) = \sum_{k=1}^{d+1} w_{jk} [1 - \exp(-\lambda_k u)], \quad u \geq 0$$

where $\lambda_k > 0$, $\sum_{k=1}^{d+1} w_{jk} \lambda_k \exp(-\lambda_k u) \geq 0$ and $\sum_{k=1}^{d+1} w_{jk} = 1$. Note that the weights $\{w_{jk}\}$, unlike for hyperexponential distributions, are not required to be all positive. Further, we suppose that $\lambda_1, \dots, \lambda_{d+1}$ are distinct and explicitly known and that the weight vectors of distinct random variables differ at least in one component. Let $A \in \{0, 1\}^{m \times N}$ denote an a priori known matrix which is 1-identifiable in the following sense.

DEFINITION 2.1. *A matrix A is k -identifiable if every set of $2k$ of its columns is linearly independent.*

Let $X \equiv (X_1, \dots, X_N)$ and $Y = AX$. For each $i \in [m]$, let $p_i := \{j \in [N] : a_{ij} = 1\}$. Further, we presume that, $\forall i \in [m]$, we have access to a sequence of IID samples $\{Y_{il}\}_{l \geq 1}$ of Y_i . Our problem then is to estimate for each X_j , its vector of weights $\mathbf{w}_j \equiv (w_{j1}, \dots, w_{jd})$ and consequently its complete distribution F_j , since $w_{j(d+1)} = 1 - \sum_{k=1}^d w_{jk}$.

Before developing the estimation procedure, we begin by making a case for the distribution model of (2.1).

3. Approximating distribution functions. Let $\mathcal{G} = \{G^1, \dots, G^N\}$ denote a finite family of arbitrary non-negative distributions. For the problem of simultaneously estimating all member of \mathcal{G} a useful strategy, as we demonstrate now, is to approximate each G^i by a GH distribution.

Recall that the CDF of a GH random variable X is given by

$$(3.1) \quad F_X(u) = \sum_{k=1}^{d+1} \alpha_k [1 - \exp(-\lambda_k u)], \quad u \geq 0,$$

where $\lambda_k > 0$, $\sum_{k=1}^{d+1} \alpha_k \lambda_k \exp(-\lambda_k u) \geq 0$ and $\sum_{k=1}^{d+1} \alpha_k = 1$. Consequently, its MGF is given by

$$(3.2) \quad M_X(t) = \sum_{k=1}^{d+1} \alpha_k \frac{\lambda_k}{\lambda_k + t}.$$

In addition to the simple algebraic form of the above quantities, the other major reason to use the GH class is that, in the sense of weak topology, it is dense in the

set of all non-negative distributions (see [2]). In fact, as the following result from [21] shows, we have much more.

THEOREM 3.1. *For $n, k \in \mathbb{Z}_{++}$, let $X_{n,k}$ be a nonnegative GH random variable with mean k/n , variance $\sigma_{n,k}^2$ and CDF $W_{n,k}$. Suppose*

1. *the function $\nu : \mathbb{Z}_{++} \rightarrow \mathbb{Z}_{++}$ satisfies $\lim_{n \rightarrow \infty} \nu(n)/n = \infty$.*
2. *there exists $0 < s < 1$ such that $\lim_{n \rightarrow \infty} n^{1+s} \sigma_{n,k}^2/k = 0$ uniformly with respect to k .*

Then given any continuous non-negative distribution function F , the following holds:

1. *the function F_n given by*

$$F_n(u) = \sum_{k=1}^{\nu(n)} \{F(k/n) - F((k-1)/n)\} W_{n,k}(u) \\ + (1 - F(\nu(n)/n)) W_{n, \nu(n)+1}(u)$$

is a GH distribution for every $n \in \mathbb{Z}_{++}$ and

2. *F_n converges uniformly to F , i.e.,*

$$\lim_{n \rightarrow \infty} \sup_{-\infty < u < \infty} |F_n(u) - F(u)| = 0.$$

Observe that, for each n , the exponential stage parameters of F_n depend only on the choice of the random variables $\{X_{n,k} : 1 \leq k \leq \nu(n) + 1\}$.

What this observation and the above result imply in relation to \mathcal{G} is that if we fix the random variables $\{X_{n,k}\}$ and let M^i denote the MGF of G^i , then for any given $\epsilon_1, \epsilon_2 > 0$ and any finite set $\tau = \{t_1, \dots, t_k\} \subset \mathbb{R}_+$, $\exists n \equiv n(\epsilon_1, \epsilon_2, \tau) \in \mathbb{Z}_{++}$ such that for each $i \in [N]$, G^i and its n^{th} GH approximation, F_n^i , are ϵ_1 -close in the sup norm and for each $j \in [k]$, $|M^i(t_j) - M_n^i(t_j)| \leq \epsilon_2$. Further, the exponential stage parameters are explicitly known and identical across the approximations F_n^1, \dots, F_n^N , which now justifies our model of (2.1).

The problem of estimating the individual members of \mathcal{G} can thus be reduced to determining the vector of weights that characterizes each approximation and hence each distribution.

4. Distribution tomography scheme. The outline for this section is as follows. For each $i \in [m]$, we use the IID samples of Y_i to estimate its MGF and subsequently build a polynomial system, say $H_i(\mathbf{x}) = 0$. We call this the elementary polynomial system (EPS). We then show that for each $i \in [m]$ and each $j \in p_i$, a close approximation of the vector \mathbf{w}_j is present in the solution set of $H_i(\mathbf{x}) = 0$, denoted $V(H_i)$. To match the weight vectors to the corresponding random variables, we make use of the fact that A is 1-identifiable.

4.1. Construction of elementary polynomial systems. Fix an arbitrary $i \in [m]$ and suppose that $|p_i| = N_i$, i.e., Y_i is a sum of N_i random variables, which for notational convenience, we relabel as X_1, \dots, X_{N_i} . Because of independence of the random variables, observe that the MGF of Y_i is well defined $\forall t \in \mathbb{R}_{++}$ and satisfies the relation

$$(4.1) \quad M_{Y_i}(t) = \prod_{j=1}^{N_i} \left\{ \sum_{k=1}^{d+1} w_{jk} \left(\frac{\lambda_k}{\lambda_k + t} \right) \right\}.$$

On simplification, after substituting $w_{j(d+1)} = 1 - \sum_{k=1}^d w_{jk}$, we get

$$(4.2) \quad \mu_i(t) = \left\{ \prod_{j=1}^{N_i} \left[\sum_{k=1}^d w_{jk} \Lambda_k(t) + \lambda_{d+1} \right] \right\},$$

where $\Lambda_k(t) = (\lambda_k - \lambda_{d+1})t/(\lambda_k + t)$ and $\mu_i(t) = M_{Y_i}(t)(\lambda_{d+1} + t)^{N_i}$.

For now, let us assume that we know $M_{Y_i}(t)$ and hence $\mu_i(t)$ exactly for every valid t . We will refer henceforth to this situation as the ideal case. Treating t as a parameter, we can then use (4.2) to define a canonical polynomial

$$(4.3) \quad f(\mathbf{x}; t) = \left\{ \prod_{j=1}^{N_i} \left[\sum_{k=1}^d x_{jk} \Lambda_k(t) + \lambda_{d+1} \right] \right\} - \mu_i(t),$$

where $\mathbf{x} \equiv (\mathbf{x}_1, \dots, \mathbf{x}_{N_i})$ with $\mathbf{x}_j \equiv (x_{j1}, \dots, x_{jd})$. As this is a multivariate map in $d \cdot N_i$ variables, we can choose an arbitrary set $\tau = \{t_1, \dots, t_{d \cdot N_i}\} \subset \mathbb{R}_{++}$ consisting of distinct numbers and define an intermediate square polynomial system

$$(4.4) \quad F_\tau(\mathbf{x}) \equiv (f_1(\mathbf{x}), \dots, f_{d \cdot N_i}(\mathbf{x})) = 0,$$

where $f_k(\mathbf{x}) \equiv f(\mathbf{x}; t_k)$.

Since (4.4) depends on choice of τ , analyzing it directly is difficult. But observe that i) the expansion of each f_n or equivalently (4.3) results in rational coefficients in t of the form $\Lambda_1^{\omega_1}(t) \cdots \Lambda_d^{\omega_d}(t) \lambda_{d+1}^{\omega_{d+1}}$, where, for each k , $\omega_k \in \mathbb{Z}_+$ and $\sum_{k=1}^{d+1} \omega_k = N_i$, and ii) the monomials that constitute each f_n are identical. This suggests that one may be able to get a simpler representation for (4.4). We do so in the following three steps, where the first two focus on simplifying (4.3).

Step1-Gather terms with common coefficients: Let $\boldsymbol{\omega}$ denote the d -dimensional vector $(\omega_1, \dots, \omega_d)$ and let $\boldsymbol{\Omega} \equiv (\boldsymbol{\omega}, \omega_{d+1})$. Also, let

$$\Delta_{d+1, N_i} := \left\{ \boldsymbol{\Omega} \in \mathbb{Z}_+^{d+1} : \sum_{k=1}^{d+1} \omega_k = N_i \right\}.$$

For a vector $\mathbf{b} \equiv (b_1, \dots, b_{N_i}) \in [d+1]^{N_i}$, let its type be denoted by $\Theta(\mathbf{b}) \equiv (\theta_1(\mathbf{b}), \dots, \theta_{d+1}(\mathbf{b}))$, where $\theta_k(\mathbf{b})$ is the count of the element k in \mathbf{b} . For every $\boldsymbol{\Omega} \in \Delta_{d+1, N_i}$, additionally define the set

$$\mathcal{B}_\Omega = \{\mathbf{b} \equiv (b_1, \dots, b_{N_i}) \in [d+1]^{N_i} : \Theta(\mathbf{b}) = \boldsymbol{\Omega}\}$$

and the polynomial $g(\mathbf{x}; \boldsymbol{\Omega}) = \sum_{\mathbf{b} \in \mathcal{B}_\Omega} \left(\prod_{j=1}^{N_i} \prod_{b_j \neq d+1} x_{jb_j} \right)$. For any $\boldsymbol{\omega} \in \mathbb{Z}_+^d$, let $\Lambda^\omega \equiv \Lambda_1^{\omega_1}(t) \cdots \Lambda_d^{\omega_d}(t)$. Then collecting terms with common coefficients in (4.3), the above notations help us rewrite it as

$$(4.5) \quad f(\mathbf{x}; t) = \sum_{\boldsymbol{\Omega} \in \Delta_{d+1, N_i}} g(\mathbf{x}; \boldsymbol{\Omega}) \Lambda^\omega \lambda_{d+1}^{\omega_{d+1}} - \mu_i(t).$$

Step2-Coefficient expansion and regrouping: Using an idea similar to partial fraction expansion for rational functions in t , the goal here is to decompose each Λ^ω into simpler terms. For each $j, k \in [d]$, let

$$\beta_{jk} := \begin{cases} \frac{\lambda_j(\lambda_k - \lambda_{d+1})}{(\lambda_j - \lambda_k)}, & j \neq k, \\ 1, & j = k. \end{cases}$$

For each $\omega \in \mathbb{Z}_+^d$, let $\mathcal{D}(\omega) := \{k \in [d] : \omega_k > 0\}$. Further, if $\mathcal{D}(\omega) \neq \emptyset$, then $\forall k \in \mathcal{D}(\omega)$ and $\forall q \in [\omega_k]$, let

$$\bar{\Delta}_{kq}(\omega) := \{\mathbf{s} \equiv (s_1, \dots, s_d) \in \mathbb{Z}_+^d : s_r = 0, \forall r \in \mathcal{D}(\omega)^c \cup \{k\}; \sum_{n=1}^d s_n = \omega_k - q\} \text{ and}$$

$$(4.6) \quad \gamma_{kq}(\omega) := \prod_{r \in \mathcal{D}(\omega)} \beta_{kr}^{\omega_r} \left\{ \sum_{\mathbf{s} \in \bar{\Delta}_{kq}(\omega)} \prod_{r \in \mathcal{D}(\omega)} \binom{\omega_r + s_r - 1}{\omega_r - 1} \beta_{rk}^{s_r} \right\}.$$

The desired decomposition is now given in the following result.

LEMMA 4.1. *If $\mathcal{D}(\omega) \neq \emptyset$, then*

$$(4.7) \quad \Lambda^\omega = \sum_{k \in \mathcal{D}(\omega)} \sum_{q=1}^{\omega_k} \gamma_{kq}(\omega) \Lambda_k^q(t)$$

for all t . Further, this expansion is unique.

Proof. See Appendix B. \square

By applying this expansion to each coefficient in (4.5) and regrouping, we have

$$(4.8) \quad f(\mathbf{x}; t) = \sum_{k=1}^d \sum_{q=1}^{N_i} h_{kq}(\mathbf{x}) \Lambda_k^q(t) \lambda_{d+1}^{N_i - q} - c(t),$$

where $c(t) = \mu_i(t) - \lambda_{d+1}^{N_i}$ and

$$(4.9) \quad h_{kq}(\mathbf{x}) = \sum_{\Omega \in \Delta_{d+1, N_i}, \omega_k \geq q} \frac{\gamma_{kq}(\omega)}{\lambda_{d+1}^{N_i - q - \omega_{d+1}}} g(\mathbf{x}; \Omega).$$

We consider a simple example to better illustrate the above two steps.

EXAMPLE 1. Suppose $N_i = 3$ and $d = 2$. In this case, clearly $\Lambda_k(t) = \frac{(\lambda_k - \lambda_3)t}{\lambda_k + t}$, $\mu_i(t) = (\lambda_3 + t)^3 M_{Y_i}(t)$ and $\mathbf{x} \equiv (x_{11}, x_{12}, x_{21}, x_{22}, x_{31}, x_{32})$. Equation (4.3) is

$$f(\mathbf{x}; t) = \prod_{j=1}^3 [x_{j1} \Lambda_1(t) + x_{j2} \Lambda_2(t) + \lambda_3] - \mu_i(t),$$

while (4.5) is

$$\begin{aligned} f(\mathbf{x}; t) = & x_{11}x_{21}x_{31}\Lambda_1^3(t) + x_{12}x_{22}x_{32}\Lambda_2^3(t) + \\ & (x_{11}x_{21} + x_{11}x_{31} + x_{21}x_{31})\Lambda_1^2(t)\lambda_3 + \\ & (x_{12}x_{22} + x_{12}x_{32} + x_{22}x_{32})\Lambda_2^2(t)\lambda_3 + \\ & (x_{11}x_{21}x_{32} + x_{11}x_{22}x_{31} + x_{12}x_{21}x_{31})\Lambda_1^2(t)\Lambda_2(t) + \\ & (x_{11}x_{22}x_{32} + x_{12}x_{21}x_{32} + x_{12}x_{22}x_{31})\Lambda_1(t)\Lambda_2^2(t) + \\ & (x_{11} + x_{21} + x_{31})\Lambda_1(t)\lambda_3^2 + (x_{12} + x_{22} + x_{32})\Lambda_2(t)\lambda_3^2 + \\ & (x_{11}x_{22} + x_{11}x_{32} + x_{12}x_{21} + x_{21}x_{32} + x_{12}x_{31} + x_{22}x_{31})\Lambda_1(t)\Lambda_2(t)\lambda_3 \\ & + \lambda_3^3 - \mu_i(t). \end{aligned}$$

Now observe that

$$\Lambda_1(t)\Lambda_2(t) = \beta_{12}\Lambda_1(t) + \beta_{21}\Lambda_2(t),$$

$$\Lambda_1^2(t)\Lambda_2(t) = \beta_{12}\Lambda_1^2(t) + \beta_{12}\beta_{21}\Lambda_1(t) + \beta_{21}^2\Lambda_2(t)$$

and

$$\Lambda_1(t)\Lambda_2^2(t) = \beta_{12}^2\Lambda_1(t) + \beta_{21}\Lambda_2^2(t) + \beta_{12}\beta_{21}\Lambda_2(t).$$

Substituting these identities, we can write $f(\mathbf{x}; t)$ in terms of (4.8) as

$$\begin{aligned} f(\mathbf{x}; t) = & h_{11}(\mathbf{x})\Lambda_1(t)\lambda_3^2 + h_{12}(\mathbf{x})\Lambda_1^2(t)\lambda_3 + h_{13}(\mathbf{x})\Lambda_1^3(t) + \\ & h_{21}(\mathbf{x})\Lambda_2(t)\lambda_3^2 + h_{22}(\mathbf{x})\Lambda_2^2(t)\lambda_3 + h_{23}(\mathbf{x})\Lambda_2^3(t) + \lambda_3^3 - \mu_i(t), \end{aligned}$$

where

$$\begin{aligned} h_{11}(\mathbf{x}) = & x_{11} + x_{21} + x_{31} + \frac{\beta_{12}\beta_{21}}{\lambda_3^2} (x_{11}x_{21}x_{32} + x_{11}x_{22}x_{31} + x_{12}x_{21}x_{31}) + \\ & \frac{\beta_{12}^2}{\lambda_3^2} (x_{11}x_{22}x_{32} + x_{12}x_{21}x_{32} + x_{12}x_{22}x_{31}) + \\ & \frac{\beta_{12}}{\lambda_3} (x_{11}x_{22} + x_{11}x_{32} + x_{12}x_{21} + x_{21}x_{32} + x_{12}x_{31} + x_{22}x_{31}), \end{aligned}$$

$$h_{12}(\mathbf{x}) = x_{11}x_{21} + x_{11}x_{31} + x_{21}x_{31} + \frac{\beta_{12}}{\lambda_3} (x_{11}x_{21}x_{32} + x_{11}x_{22}x_{31} + x_{12}x_{21}x_{31}),$$

$$h_{13}(\mathbf{x}) = x_{11}x_{21}x_{31},$$

and so on.

Step3-Eliminate dependence on τ : The advantage of (4.8) is that, apart from $c(t)$, the number of t -dependent coefficients equals $d \cdot N_i$ which is exactly the number of unknowns in the polynomial f . Further, as shown below, they are linearly independent.

LEMMA 4.2. *For $k \in [d \cdot N_i]$, let $b_k := \min\{j \in [d] : j \cdot N_i \geq k\}$. Then the matrix T_τ , where, for $j, k \in [d \cdot N_i]$*

$$(4.10) \quad (T_\tau)_{jk} = \Lambda_{b_k}^{k-(b_k-1) \cdot N_i}(t_j) \lambda_{d+1}^{b_k \cdot N_i - k},$$

is non-singular.

Proof. See Appendix A. \square

Observe that if we let $\mathbf{c}_\tau \equiv (c(t_1), \dots, c(t_{d \cdot N_i}))$, $\mathcal{E}_k(\mathbf{x}) \equiv (h_{k1}(\mathbf{x}), \dots, h_{kN_i}(\mathbf{x}))$ and $\mathcal{E}(\mathbf{x}) \equiv (\mathcal{E}_1(\mathbf{x}), \dots, \mathcal{E}_d(\mathbf{x}))$, then (4.4) can be equivalently expressed as

$$(4.11) \quad T_\tau \mathcal{E}(\mathbf{x}) - \mathbf{c}_\tau = 0.$$

Premultiplying (4.11) by $(T_\tau)^{-1}$, which now exists by Lemma 4.2, we have

$$(4.12) \quad \mathcal{E}(\mathbf{x}) - (T_\tau)^{-1} \mathbf{c}_\tau = 0.$$

Clearly, $\mathbf{w} \equiv (\mathbf{w}_1, \dots, \mathbf{w}_{N_i})$ is a root of (4.3) and hence of (4.12). This immediately implies that $(T_\tau)^{-1} \mathbf{c}_\tau = \mathcal{E}(\mathbf{w})$ and consequently (4.12) can be rewritten as

$$(4.13) \quad H_i(\mathbf{x}) \equiv \mathcal{E}(\mathbf{x}) - \mathcal{E}(\mathbf{w}) = 0.$$

Note that (4.13) is devoid of any reference to the set τ and can be arrived at using any valid τ . For this reason, we will henceforth refer to (4.13) as the EPS.

EXAMPLE 2. Let $N_i = d = 2$. Also, let $\lambda_1 = 5, \lambda_2 = 3$ and $\lambda_3 = 1$. Then the map \mathcal{E} described above is given by

$$(4.14) \quad \mathcal{E}(\mathbf{x}) = \begin{pmatrix} x_{11} + x_{21} + 5(x_{11}x_{22} + x_{12}x_{21}) \\ x_{11}x_{21} \\ x_{12} + x_{22} - 6(x_{11}x_{22} + x_{12}x_{21}) \\ x_{12}x_{22} \end{pmatrix}.$$

Two useful properties of the EPS are stated next. For $\sigma \in S_{N_i}$, let $\mathbf{x}_\sigma := (\mathbf{x}_{\sigma(1)}, \dots, \mathbf{x}_{\sigma(N_i)})$ denote a permutation of the vectors $\mathbf{x}_1, \dots, \mathbf{x}_{N_i}$. Further, for any $\mathbf{x} \in \mathbb{R}^{d \cdot N_i}$, let $\pi_\mathbf{x} := \{\mathbf{x}_\sigma : \sigma \in S_{N_i}\}$.

LEMMA 4.3. $H_i(\mathbf{x}) = H_i(\mathbf{x}_\sigma), \forall \sigma \in S_{N_i}$. That is, the map H_i is symmetric.

Proof. Observe that $H_i(\mathbf{x}) = (T_\tau)^{-1} F_\tau(\mathbf{x})$ and F_τ , as defined in (4.4), is symmetric. The result thus follows. \square

LEMMA 4.4. There exists an open dense set \mathcal{R}_i of $\mathbb{R}^{d \cdot N_i}$ such that if $\mathbf{w} \in \mathcal{R}_i$, then $|V(H_i)| = k \times N_i!$, where $k \in \mathbb{Z}_{++}$ is independent of $\mathbf{w} \in \mathcal{R}_i$. Further, each solution is non-singular.

Proof. See Appendix C. \square

We henceforth assume that $\mathbf{w} \in \mathcal{R}_i$. From the definition of the EPS in (4.13), it is obvious that $\mathbf{w} \in V(H_i)$. By Lemma 4.3, it also follows that if $\mathbf{x}^* \in V(H_i)$, then $\pi_{\mathbf{x}^*} \subset V(H_i)$. Hence, it suffices to work with

$$(4.15) \quad \mathcal{M}_i = \{\alpha \in \mathbb{C}^d : \exists \mathbf{x}^* \in V(H_i) \text{ with } \mathbf{x}_1^* = \alpha\}.$$

Clearly, $\mathcal{W}_i := \{\mathbf{w}_1, \dots, \mathbf{w}_{N_i}\} \subset \mathcal{M}_i$. A point to note here is that $\mathcal{I}_i := \mathcal{M}_i \setminus \mathcal{W}_i$ is not empty in general.

Our next objective is to develop the above theory for the case where for each $i \in [m]$, instead of the exact value of $M_{Y_i}(t)$, we have access only to the IID realizations $\{Y_{il}\}_{l \geq 1}$ of the random variable Y_i . That is, for each $k \in [N_i]$, we have to use the sample average $\hat{M}_{Y_i}(t_k; L) = \left(\sum_{l=1}^L \exp(-t_k Y_{il}) \right) / L$ for an appropriately chosen large L , $\hat{c}(t_k; L) = \hat{M}_{Y_i}(t_k) (\lambda_{d+1} + t_k)^{N_i} - \lambda_{d+1}^{N_i}$ and $\hat{\mathbf{c}}_{\tau, L} \equiv (\hat{c}(t_1; L), \dots, \hat{c}(t_{d \cdot N_i}; L))$ as substitutes for each $M_{Y_i}(t_k)$, each $c(t_k)$ and \mathbf{c}_τ respectively. But even then note that the noisy or the perturbed version of the EPS

$$(4.16) \quad \hat{H}_i(\mathbf{x}) \equiv \mathcal{E}(\mathbf{x}) - (T_\tau)^{-1}(\hat{\mathbf{c}}_{\tau, L}) = 0.$$

is always well defined. More importantly, the perturbation is only in its constant term. As in Lemma 4.3, it then follows that the map \hat{H}_i is symmetric.

Next observe that since \mathcal{R}_i is open (see Lemma 4.4), there exists a small enough $\bar{\delta}_i > 0$ such that $B(\mathbf{w}; \bar{\delta}_i) \subset \mathcal{R}_i$. Using the regularity of solutions of the EPS (see Lemma 4.4), the inverse function theorem then gives the following result.

LEMMA 4.5. Let $\delta \in (0, \bar{\delta}_i)$ be such that for any two distinct solutions in $V(H_i)$, say \mathbf{x}^* and \mathbf{y}^* , $B(\mathbf{x}^*; \delta) \cap B(\mathbf{y}^*; \delta) = \emptyset$. Then there exists an $\epsilon(\delta) > 0$ such that if $\mathbf{u} \in \mathbb{R}^{d \cdot N_i}$ and $\|\mathbf{u} - \mathcal{E}(\mathbf{w})\| < \epsilon(\delta)$, then the solution set $V(\hat{H}_i)$ of the perturbed EPS $\mathcal{E}(\mathbf{x}) - \mathbf{u} = 0$ satisfies the following:

1. All roots in $V(\hat{H}_i)$ are regular points of the map \mathcal{E} .
2. For each $\mathbf{x}^* \in V(H_i)$, there is one and only one $\mathbf{z}^* \in V(\hat{H}_i)$ such that $\|\mathbf{x}^* - \mathbf{z}^*\| < \delta$.

As a consequence, we have the following.

LEMMA 4.6. *Let δ and $\epsilon(\delta)$ be as described in Lemma 4.5. Then for tolerable failure rate $\kappa > 0$ and the chosen set τ , $\exists L_{\tau, \delta, \kappa} \in \mathbb{Z}_{++}$ such that if $L \geq L_{\tau, \delta, \kappa}$, then with probability greater than $1 - \kappa$, we have $\|(T_\tau)^{-1}\hat{\mathbf{c}}_{\tau, L} - \mathcal{E}(\mathbf{w})\| < \epsilon(\delta)$.*

Proof. Note that $\exp(-t_k Y_{il}) \in [0, 1] \forall i, l$ and k . The Hoeffding inequality (see [13]) then shows that for any $\epsilon > 0$, $\Pr\{|\hat{M}_{Y_i}(t_k; L) - M_{Y_i}(t_k)| > \epsilon\} \leq \exp(-2\epsilon^2 L)$. Since $\mathcal{E}(\mathbf{w}) = (T_\tau)^{-1}\mathbf{c}_\tau$, the result is now immediate. \square

The above two results, in simple words, state that solving (4.16) for a large enough L , with high probability, is almost as good as solving the EPS of (4.13). For $L \geq L_{\tau, \delta, \kappa}$, let $\mathcal{A}_i(\kappa)$ denote the event $\|(T_\tau)^{-1}\hat{\mathbf{c}}_{\tau, L} - \mathcal{E}(\mathbf{w})\| < \epsilon(\delta)$. Clearly, $\Pr\{\mathcal{A}_i^c(\kappa)\} \leq \kappa$. As in (4.15), let

$$(4.17) \quad \hat{\mathcal{M}}_i = \{\hat{\alpha} \in \mathbb{C}^d : \exists \mathbf{z}^* \in V(\hat{H}_i) \text{ with } \mathbf{z}_1^* = \alpha\}.$$

We are now done discussing the EPS for an arbitrary $i \in [m]$. In summary, we have managed to obtain a set $\hat{\mathcal{M}}_i$ in which a close approximation of the weight vectors of random variables X_j that add up to give Y_i are present with high probability. The next subsection takes a unified view of the solution sets $\{\hat{\mathcal{M}}_i : i \in [m]\}$ to match the weight vectors to the corresponding random variables. But before that, we redefine \mathcal{W}_i as $\{\mathbf{w}_j : j \in p_i\}$. Accordingly, $\mathcal{M}_i, \hat{\mathcal{M}}_i, V(H_i)$ and $V(\hat{H}_i)$ are also redefined using notations of Section 2.

4.2. Parameter matching using 1-identifiability. We begin by giving a physical interpretation for the 1-identifiability condition of the matrix A . For this, let $\mathcal{G}_j := \{i \in [m] : j \in p_i\}$ and $\mathcal{B}_j := [m] \setminus \mathcal{G}_j$.

LEMMA 4.7. *For a 1-identifiable matrix A , each index $j \in [N]$ satisfies*

$$\{j\} = \bigcap_{g \in \mathcal{G}_j} p_g \cap \bigcap_{b \in \mathcal{B}_j} p_b^c =: \mathcal{D}_j.$$

Proof. By definition, $j \in \mathcal{D}_j$. For converse, if $k \in \mathcal{D}_j$, $k \neq j$, then columns j and k of A are identical; contradicting its 1-identifiability condition. Thus $\{j\} = \mathcal{D}_j$. \square

An immediate result is the following.

COROLLARY 4.8. *Suppose A is a 1-identifiable matrix. If the map $u : [N] \rightarrow X$, where X is an arbitrary set, is bijective and $\forall i \in [m]$, $v_i := \{u(j) : j \in p_i\}$, then for each $j \in [N]$*

$$\{u(j)\} = \bigcap_{g \in \mathcal{G}_j} v_g \cap \bigcap_{b \in \mathcal{B}_j} v_b^c$$

By reframing this, we get the following result.

THEOREM 4.9. *Suppose A is a 1-identifiable matrix. If the weight vectors $\mathbf{w}_1, \dots, \mathbf{w}_N$ are pairwise distinct, then the rule*

$$(4.18) \quad \psi : j \rightarrow \bigcap_{g \in \mathcal{G}_j} \mathcal{W}_g \cap \bigcap_{b \in \mathcal{B}_j} \mathcal{W}_b^c,$$

satisfies $\psi(j) = \mathbf{w}_j$.

This result is where the complete potential of the 1-identifiability condition of A is being truly taken advantage of. What this states is that if we had access to the collection of sets $\{\mathcal{W}_i : i \in [m]\}$, then using ψ we would have been able to uniquely match the weight vectors to the random variables. But note that, at present, we have access only to the collection $\{\mathcal{M}_i : i \in [m]\}$ in the ideal case and $\{\hat{\mathcal{M}}_i : i \in [m]\}$ in the perturbed case. In spite of this, we now show that if $\forall i_1, i_2 \in [m], i_1 \neq i_2$,

$$(4.19) \quad \mathcal{I}_{i_1} \cap \mathcal{M}_{i_2} = \emptyset,$$

a condition that always held in simulation experiments, then the rules:

$$(4.20) \quad \psi : j \rightarrow \bigcap_{g \in \mathcal{G}_j} \mathcal{M}_g \cap \bigcap_{b \in \mathcal{B}_j} \mathcal{M}_b^c$$

for the ideal case, and

$$(4.21) \quad \hat{\psi} : j \rightarrow \bigcap_{g \in \mathcal{G}_j} \hat{\mathcal{M}}_g \cap \bigcap_{b \in \mathcal{B}_j} \hat{\mathcal{M}}_b^c$$

in the perturbed case, with minor modifications recover the correct weight vector associated to each random variable X_j .

We first discuss the ideal case. Let $\mathcal{S} := \{j \in [N] : |\mathcal{G}_j| \geq 2\}$. Because of (4.19) and Theorem 4.9, note that

1. If $j \in \mathcal{S}$, then $\psi(j) = \{\mathbf{w}_j\}$.
2. If $j \in \mathcal{S}^c$ and $j \in p_{i^*}$, then $\psi(j) = \{\mathbf{w}_j\} \cup \mathcal{I}_{i^*}$.

That is, (4.20) works perfectly fine when $j \in \mathcal{S}$. The problem arises only when $j \in \mathcal{S}^c$ as $\psi(j)$ does not give us output a unique vector. To correct this, fix $j \in \mathcal{S}^c$. If $j \in p_{i^*}$, then let $\mathbf{v}^{sub} \equiv (\mathbf{w}_k : k \in p_{i^*} \setminus \{j\})$. Because of 1-identifiability, note that if $k \in p_{i^*} \setminus \{j\}$, then $k \in \mathcal{S}$. From (4.3) and (4.4), it is also clear that $(\mathbf{v}^{sub}, \alpha) \in V(H_i)$ if and only if $\alpha = \mathbf{w}_j$. This suggests that we need to match parameters in two stages. In stage 1, we use (4.20) to assign weight vectors to all those random variables X_j such that $j \in \mathcal{S}$. In stage 2, for each $j \in \mathcal{S}^c$, we identify $i^* \in [m]$ such that $j \in p_{i^*}$. We then construct \mathbf{v}^{sub} . We then assign to j that unique α for which $(\mathbf{v}^{sub}, \alpha) \in V(H_{i^*})$. Note that we are ignoring the trivial case where $|p_{i^*}| = 1$. It is now clear that by using (4.20) with modifications as described above, at least for the ideal case, we can uniquely recover back for each random variable X_j its corresponding weight vector \mathbf{w}_j .

We next handle the case of noisy measurements. Let $\mathcal{U} := \cup_{i \in [m]} \mathcal{M}_i$ and $\hat{\mathcal{U}} := \cup_{i \in [m]} \hat{\mathcal{M}}_i$. Observe that using (4.21) directly, with probability one, will satisfy $\hat{\psi}(j) = \emptyset$ for each $j \in [N]$. This happens because we are distinguishing across the solution sets the estimates obtained for a particular weight vector. Hence as a first step we need to define a relation \sim on $\hat{\mathcal{U}}$ that associates these related elements. Recall from Lemmas 4.5 and 4.6 that the set $\hat{\mathcal{M}}_i$ can be constructed for any small enough choice of $\delta, \kappa > 0$. With choice of δ that satisfies

$$(4.22) \quad 0 < 4\delta < \min_{\alpha, \beta \in \mathcal{U}} \|\alpha - \beta\|,$$

let us consider the event $\mathcal{A} := \cap_{i \in [m]} \mathcal{A}_i(\kappa/m)$. Using a simple union bound, it follows that $\Pr\{\mathcal{A}^c\} \leq \kappa$. Now suppose that the event \mathcal{A} is a success. Then by (4.22) and Lemma 4.5, the following observations follow trivially.

1. For each $i \in [m]$ and each $\alpha \in \mathcal{M}_i$, there exists at least one $\hat{\alpha} \in \hat{\mathcal{M}}_i$ such that $\|\hat{\alpha} - \alpha\| < \delta$.
2. For each $i \in [m]$ and each $\hat{\alpha} \in \hat{\mathcal{M}}_i$, there exists precisely one $\alpha \in \mathcal{M}_i$ such that $\|\hat{\alpha} - \alpha\| < \delta$.
3. Suppose for distinct elements $\alpha, \beta \in \mathcal{U}$, we have $\hat{\alpha}, \hat{\beta} \in \hat{\mathcal{U}}$ such that $\|\hat{\alpha} - \alpha\| < \delta$ and $\|\hat{\beta} - \beta\| < \delta$. Then $\|\hat{\alpha} - \hat{\beta}\| > 2\delta$.

From these, it is clear that the relation \sim on $\hat{\mathcal{U}}$ should be

$$(4.23) \quad \hat{\alpha} \sim \hat{\beta} \text{ iff } \|\hat{\alpha} - \hat{\beta}\| < 2\delta.$$

It is also easy to see that, whenever the event \mathcal{A} is a success, \sim defines an equivalence relation on $\hat{\mathcal{U}}$. For each $i \in [m]$, the obvious idea then is to replace each element of $\hat{\mathcal{M}}_i$ and its corresponding d -dimensional component in $V(\hat{H}_i)$ with its equivalence class. It now follows that (4.21), with modifications as was done for the ideal case, will satisfy

$$(4.24) \quad \hat{\psi}(j) = \{\hat{\alpha} \in \hat{\mathcal{U}} : \|\hat{\alpha} - \mathbf{w}_j\| < \delta\}.$$

This is obviously the best we could have done starting from the set $\{\hat{\mathcal{M}}_i : i \in [m]\}$.

We end this section by summarizing our complete method in an algorithmic fashion.

ALGORITHM 1. *Distribution tomography*

Phase 1: Construct \mathcal{E} & Solve the EPS.

For each $i \in [m]$,

1. *Choose an arbitrary $\tau = \{t_1, \dots, t_{d \cdot N_i}\}$ of distinct positive real numbers.*
2. *Pick a large enough $L \in \mathbb{Z}_{++}$. Set $\hat{M}_{Y_i}(t_j) = \left(\sum_{l=1}^L \exp(-t_j Y_{il}) \right) / L$, $\hat{\mu}_i(t_j) = (\lambda_{d+1} + t_j)^{N_i} \hat{M}_{Y_i}(t_j)$ and $\hat{c}(t_j) = \hat{\mu}_i(t_j) - \lambda_{d+1}^{N_i}$ for each $j \in [N_i]$. Using this, construct $\hat{c}_\tau \equiv (\hat{c}(t_1), \dots, \hat{c}(t_{d \cdot N_i}))$.*
3. *Solve $\mathcal{E}(\mathbf{x}) - T_\tau^{-1} \hat{c}_\tau = 0$ using any standard solver for polynomial systems.*
4. *Build $\hat{\mathcal{M}}_i = \{\alpha \in \mathbb{C}^d : \exists \mathbf{x}^* \in V(\hat{H}_i) \text{ with } \mathbf{x}_1^* = \alpha\}$.*

Phase 2: Parameter Matching

1. *Set $\hat{\mathcal{U}} := \bigcup_{i \in [m]} \hat{\mathcal{M}}_i$. Choose $\delta > 0$ small enough and define the relation \sim on $\hat{\mathcal{U}}$, where $\hat{\alpha} \sim \hat{\beta}$ if and only if $\|\hat{\alpha} - \hat{\beta}\|_2 < 2\delta$. If \sim is not an equivalence relation, then choose a smaller δ and repeat.*
2. *Construct the quotient set $\hat{\mathcal{U}} / \sim$. Replace all elements of each $\hat{\mathcal{M}}_i$ and each $V(\hat{H}_i)$ with their equivalence class.*
3. *For each $j \in \mathcal{S}$, set $\hat{\psi}(j) = \left(\bigcap_{g \in \mathcal{G}_j} \hat{\mathcal{M}}_g \right) \cap \left(\bigcap_{b \in \mathcal{B}_j} \hat{\mathcal{M}}_b^c \right)$.*
4. *For each $j \in \mathcal{S}^c$,*
 - (a) *Set $i^* = i \in [m]$ such that $j \in p_{i^*}$.*
 - (b) *Construct $\mathbf{v}^{sub} \equiv (\hat{\psi}(k) : k \in p_{i^*}, k \neq j)$.*
 - (c) *Set $\hat{\psi}(j) = \hat{\alpha}$ such that $(\mathbf{v}^{sub}, \hat{\alpha}) \in V(\hat{H}_{i^*})$.*

5. Universality. The crucial step in the DT scheme described above was to come up with, for each $i \in [m]$, a well behaved polynomial system, i.e., one that satisfies the properties of Lemma 4.4, based solely on the samples of Y_i . Once that was done, the ability to match parameters to the component random variables was only a consequence of the 1-identifiability condition of the matrix A . This suggests that it may be possible to develop similar schemes even in settings different to the ones assumed in Section 2. In fact, functions other than the MGF could also serve as

blueprints for constructing the polynomial system. We discuss in brief few of these ideas in this section. Note that we are making a preference for polynomial systems for the sole reason that there exist computationally efficient algorithms, see for example [24, 16, 19, 28], to determine all its roots.

Consider the case, where $\forall j \in [N]$, the distribution of X_j is the finite mixture model

$$(5.1) \quad F_j(u) = \sum_{k=1}^{d_j+1} w_{jk} \phi_{jk}(u),$$

where $d_j \in \mathbb{Z}_{++}$, $w_{j1}, \dots, w_{j(d_j+1)}$ denote the mixing weights, i.e., $w_{jk} \geq 0$ and $\sum_{k=1}^{d_j+1} w_{jk} = 1$, and $\{\phi_{jk}(u)\}$ are some basis functions, say Gaussian, uniform, etc. The MGF of each X_j is clearly given by

$$(5.2) \quad M_{X_j}(t) = \sum_{k=1}^{d_j+1} w_{jk} \int_{u=0}^{\infty} \exp(-ut) d\phi_{jk}(u).$$

Now note that if the basis functions $\{\phi_{jk}\}$ are completely known, then the MGF of each Y_i will again be a polynomial in the mixing weights, $\{w_{jk}\}$, similar in spirit to the relation of (4.1). As a result, the complete recipe of Section 4 can again be attempted to estimate the weight vectors of the random variables X_j using only the IID samples of each Y_i .

In relation to (2.1) or (5.1), observe next that $\forall n \in \mathbb{Z}_{++}$, the n^{th} moment of each X_j is given by

$$(5.3) \quad \mathbb{E}(X_j^n) = \sum_{k=1}^{d_j+1} w_{jk} \int_{u=0}^{\infty} u^n d\phi_{jk}(u).$$

Hence, the n^{th} moment of Y_i is again a polynomial in the unknown weights. This suggests that, instead of the MGF, one could use the estimates of the moments of Y_i to come up with an alternative polynomial system and consequently solve for the distribution of each X_j .

Moving away from the models of (2.1) and (5.1), suppose that for each $j \in [N]$, $X_j \sim \exp(m_j)$. Assume that each mean $m_j < \infty$ and that $m_{j_1} \neq m_{j_2}$ when $j_1 \neq j_2$. We claim that the basic idea of our method can be used here to estimate m_1, \dots, m_N and hence the complete distribution of each X_j using only the samples of Y_i . As the steps are quite similar when either i) we know $M_{Y_i}(t)$ for each $i \in [m]$ and every valid t and ii) we have access only to the IID samples $\{Y_{il}\}_{l \geq 1}$ for each $i \in [m]$, we take up only the first case.

Fix $i \in [m]$ and let $p_i := \{j \in [N] : a_{ij} = 1\}$. To simplify notations, let us relabel the random variables $\{X_j : j \in p_i\}$ that add up to give Y_i as X_1, \dots, X_{N_i} , where $N_i = |p_i|$. Observe that the MGF of Y_i , after inversion, satisfies

$$(5.4) \quad \prod_{j=1}^{N_i} (1 + tm_j) = 1/M_{Y_i}(t).$$

Using (5.4), we can then define the canonical polynomial

$$(5.5) \quad f(\mathbf{x}; t) := \prod_{j=1}^{N_i} (1 + tx_j) - c(t),$$

where $\mathbf{x} \equiv (x_1, \dots, x_{N_i})$ and $c(t) = 1/M_{Y_i}(t)$. Now choose an arbitrary set $\tau = \{t_1, \dots, t_{N_i}\} \subset \mathbb{R}_{++}$ consisting of distinct numbers and define

$$(5.6) \quad F_\tau(\mathbf{x}) \equiv (f_1(\mathbf{x}), \dots, f_{N_i}(\mathbf{x})) = 0,$$

where $f_k(\mathbf{x}) = f(\mathbf{x}; t_k)$. We emphasize that this system is square of size N_i , depends on the choice of subset τ and each polynomial f_k is symmetric with respect to the variables x_1, \dots, x_{N_i} . In fact, if we let $\mathbf{c}_\tau \equiv (c(t_1), \dots, c(t_{N_i}))$ and $\mathcal{E}(\mathbf{x}) \equiv (e_1(\mathbf{x}), \dots, e_{N_i}(\mathbf{x}))$, where $e_k(\mathbf{x}) = \sum_{1 \leq j_1 < j_2 < \dots < j_k \leq N_i} x_{j_1} \cdots x_{j_k}$ denotes the k^{th} elementary symmetric polynomial in the N_i variables x_1, \dots, x_{N_i} , we can rewrite (5.6) as

$$(5.7) \quad T_\tau \mathcal{E}(\mathbf{x}) - (\mathbf{c}_\tau - \mathbf{1}) = 0.$$

Here T_τ denotes a Vandermonde matrix of order N_i in t_1, \dots, t_{N_i} with $(T_\tau)_{jk} = t_j^k$. Its determinant, given by $\det(T_\tau) = \left(\prod_{j=1}^{N_i} t_j \right) \prod_{j>i} (t_j - t_i)$, is clearly non-zero. Premultiplying (5.7) by T_τ^{-1} , we have

$$(5.8) \quad \mathcal{E}(\mathbf{x}) - T_\tau^{-1}(\mathbf{c}_\tau - \mathbf{1}) = 0.$$

Observe now that the vector $\mathbf{m} \equiv (m_1, \dots, m_{N_i})$ is a natural root of (5.6) and hence of (5.8). Hence $T_\tau^{-1}(\mathbf{c}_\tau - \mathbf{1}) = \mathcal{E}(\mathbf{m})$. The EPS for this case can thus be written as

$$(5.9) \quad H_i(\mathbf{x}) \equiv \mathcal{E}(\mathbf{x}) - \mathcal{E}(\mathbf{m}) = 0.$$

We next discuss the properties of this EPS, or more specifically, its solution set. For this, let $V(H_i) := \{\mathbf{x} \in \mathbb{C}^{N_i} : H_i(\mathbf{x}) = 0\}$.

LEMMA 5.1. $V(H_i) = \pi_{\mathbf{m}} := \{\sigma(\mathbf{m}) : \sigma \in S_{N_i}\}$.

Proof. This follows directly from (5.9). \square

LEMMA 5.2. For every $\mathbf{x}^* \in V(H_i)$, $\det(\dot{\mathcal{E}}(\mathbf{x}^*)) \neq 0$.

Proof. This follows from the fact that $\det(\dot{\mathcal{E}}(\mathbf{x})) = \prod_{1 \leq j < k \leq N_i} (x_j - x_k)$. \square

Because of Lemma 5.1, it suffices to work with only the first components of the roots. Hence we define

$$(5.10) \quad \mathcal{M}_i := \{\alpha^* \in \mathbb{C} : \exists \mathbf{x}^* \in V(H_i) \text{ with } \mathbf{x}_1^* = \alpha^*\},$$

which in this case is equivalent to the set $\{m_1, \dots, m_{N_i}\}$. Reverting back to global notations, note that

$$(5.11) \quad \mathcal{M}_i = \{m_j : j \in p_i\}.$$

Since i was arbitrary, we can repeat the above procedure to obtain the collection of solution sets $\{\mathcal{M}_i : i \in [m]\}$. Arguing as in Theorem 4.9, it is now follows that if A is 1-identifiable, then the rule

$$(5.12) \quad \psi(j) = \bigcap_{g \in \mathcal{G}_j} \mathcal{M}_g \cap \bigcap_{b \in \mathcal{B}_j} \mathcal{M}_b^c,$$

where $\mathcal{G}_j = \{i \in [m] : j \in p_i\}$ and $\mathcal{B}_j = [m] \setminus \mathcal{G}_j$, satisfies the relation $\psi(j) = m_j$. That is, having obtained the sets $\{\mathcal{M}_i : i \in [m]\}$, one can use ψ to match the parameters to the corresponding random variables.

This clearly demonstrates that even if a transformation of the MGF is a polynomial in the parameters to be estimated, our method may be applicable.

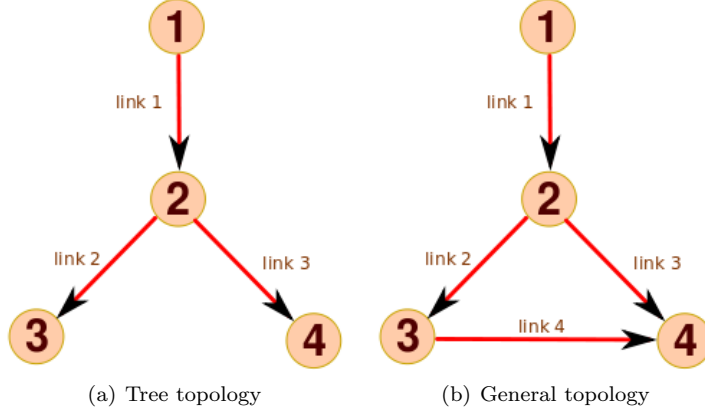


FIG. 6.1. Networks for simulation experiments.

6. Experimental Results. We assess the performance of our DT scheme using matlab based simulation experiments.

We consider the simplified network delay tomography setup wherein, given a sequence of end-to-end measurements of delay a probe packets experiences across a subset of paths in a network, we are required to estimate the delay distribution across each link. In particular, we suppose that the topology of the network is known a priori in the form of its path-link matrix, denoted $A \in \{0, 1\}^{m \times N}$, and is unvarying during the measurement phase. The rows of A correspond to the paths across which probe packets can be transmitted and its delay measured, while the columns correspond to individual links. Further, the element a_{ij} is 1 precisely when the j^{th} link is present on the i^{th} path. We let X_j , a GH random variable, denote the probe packet delay across link j and Y_i the delay across path i . We assume that the delay across different links are independent. If we let $X \equiv (X_1, \dots, X_N)$ and $Y \equiv (Y_1, \dots, Y_m)$, then observe that $Y = AX$. Clearly, this setup now resembles the model of Section 2.

We simulate the networks given in Figure 6. The specifics of each experiment and observations made are described next. Note that, unless specified otherwise, all values are rounded to 2 significant digits.

6.1. Network with Tree Topology. We work here with the network of Figure 6.1(a). Node 1 is the source node, while nodes 3 and 4 act as sink. Path p_1 connects the nodes 1, 2 and 3, while path p_2 connects the nodes 1, 2 and 4. The path-link matrix is thus given by

$$A = \begin{pmatrix} 1 & 1 & 0 \\ 1 & 0 & 1 \end{pmatrix}.$$

EXPERIMENT 1. We set the count of exponential stages in each link distribution to three. That is, we set $d = 2$. We take the corresponding exponential stage parameters λ_1, λ_2 and λ_3 to be 5, 3 and 1 respectively. For each link, the weight associated with each exponential stage is set as given in the columns labeled w_{j1}, \dots, w_{j3} of Table 6.1.

We first focus on path p_1 . Observe that its EPS is given by the map of (4.14). We collect now a million samples of its end-to-end delay. Choosing an arbitrary set $\tau = \{1.9857, 2.3782, 0.3581, 8.8619\}$, we run the first phase of Algorithm 1 to obtain $\hat{\mathcal{M}}_1$. This set along with its ideal counterpart is given in Table 6.2.

TABLE 6.1
Actual and Estimated weights for Expt. 1

Link	w_{j1}	w_{j2}	w_{j3}	\hat{w}_{j1}	\hat{w}_{j2}	\hat{w}_{j3}
1	0.17	0.80	0.03	0.15	0.82	0.02
2	0.13	0.47	0.40	0.15	0.46	0.39
3	0.80	0.15	0.05	0.79	0.15	0.06

TABLE 6.2
Solution set of the EPS for path p_1 in Expt. 1

sol-ID	\mathcal{M}_1	$\hat{\mathcal{M}}_1$
1	(0.1300, 0.4700)	(0.1542, 0.4558)
2	(0.1700, 0.8000)	(0.1292, 0.8356)
3	(3.8304, -2.8410)	(3.8525, -2.8646)
4	(0.1933, 0.7768)	(0.2260, 0.7394)
5	(0.1143, 0.4840)	(0.0882, 0.5152)
6	(0.0058, -0.1323)	(0.0052, -0.1330)

Similarly, by probing the path p_2 with another million samples and with $\tau = \{0.0842, 0.0870, 0.0305, 0.0344\}$, we determine $\hat{\mathcal{M}}_2$. The sets \mathcal{M}_2 and $\hat{\mathcal{M}}_2$ are given in Table 6.3.

TABLE 6.3
Solution set of the EPS for path p_2 in Expt. 1

sol-ID	\mathcal{M}_2	$\hat{\mathcal{M}}_2$
1	(0.8000, 0.1500)	(0.7933, 0.1459)
2	(0.1660, 0.7775)	(0.1720, 0.8095)
3	(5.5623, -4.5638)	(5.5573, -4.5584)
4	(0.1700, 0.8000)	(0.1645, 0.7669)
5	(0.8191, 0.1543)	(0.8296, 0.1540)
6	(0.0245, -0.0263)	(0.0246, -0.0259)

To match the weight vectors to corresponding links, firstly observe that the minimum distance between $\hat{\mathcal{M}}_1$ and $\hat{\mathcal{M}}_2$ is 0.0502. Based on this, we choose $\delta = 0.03$ and run the second phase of Algorithm 1. The obtained results are given in the second half of Table 6.1. Note that the weights obtained for the first link are determined by taking a simple average of the solutions obtained from the two different paths. The norm of the error vector is 0.0443.

EXPERIMENT 2. Keeping other things unchanged as in the setup of experiment 1, we consider here four exponential stages in the distribution of each X_j . The exponential stage parameters $\lambda_1, \lambda_2, \lambda_3$ and λ_4 equal 5, 4, 0.005 and 1 respectively. The corresponding weights are given in Table 6.4. But observe that the weights of the third stage is negligible for all three links. Because of this, we ignore its presence completely. That is, we consider $d = 2$ and then run Algorithm 1. The results obtained are given in the second half of Table 6.4. The norm of the error vector is 0.1843.

This clearly demonstrates that the exponential stages, which are insignificant across all link distributions, can be ignored.

TABLE 6.4
Actual and Estimated weights for Expt. 2

Link	w_{j1}	w_{j2}	w_{j3}	w_{j4}	\hat{w}_{j1}	\hat{w}_{j2}	\hat{w}_{j3}	\hat{w}_{j4}
1	0.71	0.20	0.0010	0.08	0.77	0.20	0	0.03
2	0.41	0.17	0.0015	0.41	0.38	0.18	0	0.44
3	0.15	0.80	0.0002	0.04	0.12	0.70	0	0.18

6.2. Network with General Topology. We deal here with the network of Figure 6.1(b). Nodes 1 and 2 act as source while nodes 3 and 4 act as sink. We consider here three paths. Path p_1 connects the nodes 1, 2 and 3, path p_2 connects the nodes 1, 2 and 4, while path p_3 connects the nodes 2, 3 and 4. The path-link matrix is

$$A = \begin{pmatrix} 1 & 1 & 0 & 0 \\ 1 & 0 & 1 & 0 \\ 0 & 1 & 0 & 1 \end{pmatrix}.$$

EXPERIMENT 3. The values of $d, \lambda_1, \lambda_2, \lambda_3$ are set as in Experiment 1. By choosing again a million probe packets for each path, we run Algorithm 1. The actual and estimated weights are shown in Table 6.5.

TABLE 6.5
Actual and Estimated weights for Expt. 3

Link	w_{j1}	w_{j2}	w_{j3}	\hat{w}_{j1}	\hat{w}_{j2}	\hat{w}_{j3}
1	0.34	0.26	0.40	0.34	0.24	0.42
2	0.46	0.49	0.05	0.45	0.50	0.05
3	0.12	0.65	0.23	0.11	0.68	0.21
4	0.71	0.19	0.10	0.69	0.18	0.13

The ease with which our algorithm can handle even networks that have non-tree topologies is clearly demonstrated in this experiment.

7. Discussion. This paper took advantage of the properties of polynomial systems to develop a novel algorithm for the GNT problem. For an arbitrary matrix A , which is 1-identifiable, it demonstrated successfully how to accurately estimate the distribution of the random vector X , with mutually independent components, using only IID samples of the components of the random vector $Y = AX$. Translating to network terminology, this means that one can now address the tomography problem even for networks with arbitrary topologies using only pure unicast probe packet measurements. The fact that we need only the IID samples of the components of Y shows that the processes to acquire these samples across different paths can be asynchronous. Another nice feature of this approach is that it can estimate the unknown link level performance parameters even when no prior information is available about the same.

Appendix A. Nonsingularity of coefficient matrix. In this section, we prove three results which will together demonstrate the validity of Lemma 4.2.

LEMMA A.1. Let $\omega_1, \omega_2 \in \mathbb{Z}_{++}$. Further suppose that λ_1, λ_2 and $t_1, \dots, t_{\omega_1 + \omega_2}$ are strictly positive real numbers satisfying $\lambda_1 \neq \lambda_2$ and $t_{k_1} \neq t_{k_2}$, if $k_1 \neq k_2$. Then

the square matrix

$$W = \left(\begin{array}{ccc|ccc} \frac{1}{(\lambda_1+t_1)^{\omega_1}} & \cdots & \frac{1}{(\lambda_1+t_1)} & \frac{1}{(\lambda_2+t_1)^{\omega_2}} & \cdots & \frac{1}{(\lambda_2+t_2)} \\ \vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\ \frac{1}{(\lambda_1+t_{\omega_1+\omega_2})^{\omega_1}} & \cdots & \frac{1}{(\lambda_1+t_{\omega_1+\omega_2})} & \frac{1}{(\lambda_2+t_{\omega_1+\omega_2})^{\omega_2}} & \cdots & \frac{1}{(\lambda_2+t_{\omega_1+\omega_2})} \end{array} \right)$$

is non-singular.

Proof. It suffices to show that

$$(A.1) \quad \det(W) = \left(\prod_{j=1}^{\omega_1+\omega_2} \prod_{k=1}^2 \frac{(\lambda_2 - \lambda_1)^{\omega_1\omega_2}}{(\lambda_k + t_j)^{\omega_k}} \right) \left(\prod_{1 \leq j_1 < j_2 \leq (\omega_1+\omega_2)} (t_{j_2} - t_{j_1}) \right).$$

As a first step, we perform on W the row operations

$$\text{row}_j \leftarrow \text{row}_j \times (\lambda_1 + t_j)^{\omega_1} (\lambda_2 + t_j)^{\omega_2}$$

for each $j \in [\omega_1 + \omega_2]$ to get the matrix B . Note that for $j, k \in [\omega_1 + \omega_2]$,

$$(A.2) \quad (B)_{jk} = \begin{cases} (\lambda_1 + t_j)^{k-1} (\lambda_2 + t_j)^{\omega_2}, & k \in [\omega_1] \\ (\lambda_1 + t_j)^{\omega_1} (\lambda_2 + t_j)^{k-1-\omega_1}, & \text{otherwise.} \end{cases}$$

From the properties of determinants, it follows that

$$\det(W) = \left(\prod_{j=1}^{\omega_1+\omega_2} \frac{1}{(\lambda_1 + t_j)^{\omega_1} (\lambda_2 + t_j)^{\omega_2}} \right) \det(B).$$

To verify (A.1), it only remains to show that

$$(A.3) \quad \det(B) = (\lambda_2 - \lambda_1)^{\omega_1\omega_2} \prod_{1 \leq j_1 < j_2 \leq (\omega_1+\omega_2)} (t_{j_2} - t_{j_1}).$$

Towards this, our approach is to treat λ_2 as a variable and the other indeterminates, i.e., $\lambda_1, t_1, \dots, t_{\omega_1+\omega_2}$ as constants and show that $g(\lambda_2)$ is a univariate polynomial of degree $\omega_1\omega_2$ with λ_1 as its sole root with multiplicity $\omega_1\omega_2$.

We now introduce some notations. Let $b_k(\lambda_2)$ denote the k^{th} column of B and $b_k^{r_k}(\lambda_2)$ its element-wise r_k^{th} derivative with respect to λ_2 . We use $g(\lambda_2)$ and $g^n(\lambda_2)$ to refer respectively to $\det(B)$ and its n^{th} derivative with respect to λ_2 . Lastly, for $\omega, n \in \mathbb{Z}_{++}$, we use $\Delta_{\omega,n}$ to represent the set of non-negative integer valued vector solutions of the equation $\alpha_1 + \dots + \alpha_\omega = n$.

We now prove (A.3) by equivalently showing the following:

- i. $g^n(\lambda_2)|_{\lambda_2=\lambda_1} = 0$ for each $0 \leq n < \omega_1\omega_2$.
- ii. $g^{\omega_1\omega_2}(\lambda_2)|_{\lambda_2=\lambda_1} = (\omega_1\omega_2)! \prod_{1 \leq j_1 < j_2 \leq (\omega_1+\omega_2)} (t_{j_2} - t_{j_1}) \neq 0$.
- iii. For all $n > \omega_1\omega_2$, $g^n(\lambda_2) \equiv 0$.

Firstly note that, for any $n \in \mathbb{Z}_{++}$,

$$(A.4) \quad g^n(\lambda_2) = \sum_{\Delta_{\omega_1+\omega_2,n}} \frac{n!}{r_1! \dots r_{\omega_1+\omega_2}!} \det \left(\begin{bmatrix} b_1^{r_1}(\lambda_2) & \cdots & b_{\omega_1+\omega_2}^{r_{\omega_1+\omega_2}}(\lambda_2) \end{bmatrix} \right).$$

Our strategy is to deal with $\det \left(\begin{bmatrix} b_1^{r_1}(\lambda_2) & \cdots & b_{\omega_1+\omega_2}^{r_{\omega_1+\omega_2}}(\lambda_2) \end{bmatrix} \right) =: \mu(\lambda_2; \mathbf{r})$ for every possible pattern of the tuple $\mathbf{r} \equiv (r_1, \dots, r_{\omega_1+\omega_2}) \in \mathbb{Z}_+^{\omega_1+\omega_2}$. Although the different possibilities are huge in number, the following observations, which follow directly from (A.2), will reveal that in almost all combinations of \mathbf{r} either a column becomes itself zero or is a scalar multiple of another column. In fact there is only one unique pattern where the determinant will have to be actually evaluated.

1. The column $b_{\omega_1+1}(\lambda_2)$ is a constant with respect to λ_2 . This implies that if $r_{\omega_1+1} \geq 1$, then $b_{\omega_1+1}^{r_{\omega_1+1}}(\lambda_2) \equiv \mathbf{0}$ and consequently $\mu(\lambda_2; \mathbf{r}) \equiv 0$ irrespective of what values $r_k, k \neq \omega_1+1$ take. Hence we need to focus on only those tuples \mathbf{r} where $r_{\omega_1+1} = 0$.
2. Suppose that $2 \leq k \leq \omega_2$ and $r_{\omega_1+1}, r_{\omega_1+2}, \dots, r_{\omega_1+k-1} = 0$. Then

$$b_{\omega_1+k}^{r_{\omega_1+k}}(\lambda_2) \equiv \begin{cases} \mathbf{0}, & r_{\omega_1+k} \geq k \\ \left(\prod_{j=1}^{r_{\omega_1+k}} (k-j) \right) b_{\omega_1+k-r_{\omega_1+k}}(\lambda_2), & 1 \leq r_{\omega_1+k} \leq k-1. \end{cases}$$

In other words, if $r_{\omega_1+1}, r_{\omega_1+2}, \dots, r_{\omega_1+k-1} = 0$ and $r_{\omega_1+k} > 0$, then $\mu(\lambda_2; \mathbf{r}) \equiv 0$. An inductive argument on k then shows that we need to deal with only those tuples where $r_{\omega_1+1} = \dots = r_{\omega_1+\omega_2} = 0$. For the next 2 observations, we assume that this condition explicitly holds for the tuple under consideration.

3. For column ω_1 note that if $r_{\omega_1} \geq \omega_2 + 1$, then $b_{\omega_1}^{r_{\omega_1}}(\lambda_2) \equiv \mathbf{0}$. But when $r_{\omega_1} = \omega_2$, then $b_{\omega_1}^{r_{\omega_1}}(\lambda_2) \equiv \omega_2! \left((\lambda_1 + t_1)^{\omega_1-1}, \dots, (\lambda_1 + t_{\omega_1+\omega_2})^{\omega_1-1} \right)$. On the other hand, when $\lambda_2 = \lambda_1$,

$$b_{\omega_1}^{r_{\omega_1}}(\lambda_2) \Big|_{\lambda_2=\lambda_1} = \begin{cases} b_{\omega_1+\omega_2}(\lambda_2) \Big|_{\lambda_2=\lambda_1}, & r_{\omega_1} = 0 \\ \left(\prod_{j=0}^{r_{\omega_1}-1} (\omega_2 - j) \right) b_{\omega_1+\omega_2-r_{\omega_1}}(\lambda_2) \Big|_{\lambda_2=\lambda_1}, & 1 \leq r_{\omega_1} \leq \omega_2 - 1. \end{cases}$$

This implies that $\mu(\lambda_2; \mathbf{r}) \equiv 0$ if $r_{\omega_1} \geq \omega_2 + 1$, while $\mu(\lambda_2; \mathbf{r}) \Big|_{\lambda_2=\lambda_1} = 0$ if $0 \leq r_{\omega_1} \leq \omega_2 - 1$. For our purposes, it thus remains to investigate only those tuples where $r_{\omega_1} = \omega_2$.

4. The following holds whenever $\omega_1 > 1$. Suppose that there exists $k \in [\omega_1 - 1]$ such that $r_{\omega_1} = r_{\omega_1-1} = \dots = r_{\omega_1-k+1} = \omega_2$. It is then immediate that $b_{\omega_1-k}^{r_{\omega_1-k}}(\lambda_2) \equiv \omega_2! \left((\lambda_1 + t_1)^{\omega_1-1-k}, \dots, (\lambda_1 + t_{\omega_1+\omega_2})^{\omega_1-1-k} \right)$ if $r_{\omega_1-k} = \omega_2$, while $b_{\omega_1-k}^{r_{\omega_1-k}}(\lambda_2) \equiv 0$ if $r_{\omega_1-k} \geq \omega_2 + 1$. For the case when $r_{\omega_1-k} \in \{0\} \cup [\omega_2 - 1]$, consider the following subcases.

(a) $\omega_1 \leq \omega_2$: Here note that

$$b_{\omega_1-k}^{r_{\omega_1-k}}(\lambda_2) \Big|_{\lambda_2=\lambda_1} = \begin{cases} b_{\omega_1+\omega_2-k}(\lambda_2) \Big|_{\lambda_2=\lambda_1}, & r_{\omega_1-k} = 0 \\ \left(\prod_{j=0}^{r_{\omega_1-k}-1} (\omega_2 - j) \right) \times b_{\omega_1+\omega_2-k-r_{\omega_1-k}}(\lambda_2) \Big|_{\lambda_2=\lambda_1}, & 1 \leq r_{\omega_1-k} < \omega_2 - k \\ \frac{b_{\omega_1+\omega_2-k-r_{\omega_1-k}}^{r_{\omega_1-k}}(\lambda_2) \Big|_{\lambda_2=\lambda_1}}{(\omega_2 - r_{\omega_1-k})!}, & \omega_2 - k \leq r_{\omega_1-k} < \omega_2. \end{cases}$$

- (b) $\omega_1 > \omega_2$: If $k < \omega_2$, then $b_{\omega_1-k}^{r_{\omega_1-k}}$ behaves precisely as in the subcase above. On the other hand, when $\omega_2 \leq k \leq \omega_1 - 1$ then

$$b_{\omega_1-k}^{r_{\omega_1-k}}(\lambda_2)|_{\lambda_2=\lambda_1} = \frac{b_{\omega_1+\omega_2-k-r_{\omega_1-k}}^{\omega_2}(\lambda_2)|_{\lambda_2=\lambda_1}}{(\omega_2 - r_{\omega_1-k})!}, \quad 0 \leq r_{\omega_1-k} < \omega_2.$$

As a consequence of a simple induction on k , it now follows that $\mu(\lambda_2; \mathbf{r})|_{\lambda_2=\lambda_1} = 0$ for all tuples \mathbf{r} , where $0 \leq r_1 < \omega_2$ and for each $2 \leq i \leq \omega_1$, $r_i \in \{0\} \cup [\omega_2]$. In contrast, even if one amongst r_1, \dots, r_{ω_1} has value strictly bigger than ω_2 , then $\mu(\lambda_2; \mathbf{r}) \equiv 0$.

5. The above observations essentially show that \mathbf{r}^* , where $r_1^* = \dots = r_{\omega_1}^* = \omega_2$ and $r_{\omega_1+1}^* = \dots = r_{\omega_1+\omega_2}^* = 0$, is the only tuple for which $\mu(\lambda_2; \mathbf{r}^*)$ is required to be explicitly evaluated. Using properties of Vandermonde matrices, a simple calculation show that

$$\mu(\lambda_2; \mathbf{r}^*)|_{\lambda_2=\lambda_1} = (\omega_2!)^{\omega_1} \prod_{1 \leq k_1 < k_2 \leq \omega_1+\omega_2} (t_{k_2} - t_{k_1}) \neq 0.$$

Let $\mathcal{A}_{\leq, \mathbf{0}} = \{\mathbf{r} \equiv (\mathbf{s}, \mathbf{0}) \in \mathbb{Z}_+^{\omega_1} \times \mathbb{Z}_+^{\omega_2} : 0 \leq s_1 < \omega_2 \text{ and } 0 \leq s_k \leq \omega_2 \ \forall 2 \leq k \leq \omega_1\}$, $\mathcal{A}_{\mathbf{r}^*} = \{\mathbf{r}^*\}$ and $\mathcal{A}_{>} = \mathbb{Z}_+^{\omega_1+\omega_2} \setminus \{\mathcal{A}_{\mathbf{r}^*} \cup \mathcal{A}_{\leq, \mathbf{0}}\}$ be a partition of $\mathbb{Z}_+^{\omega_1+\omega_2}$. Then a short summary of the above observations is that

$$\mu(\lambda_2; \mathbf{r})|_{\lambda_2=\lambda_1} = \begin{cases} 0, & \mathbf{r} \in \mathcal{A}_{\leq, \mathbf{0}} \\ \mu(\lambda_2; \mathbf{r}^*)|_{\lambda_2=\lambda_1}, & \mathbf{r} \in \mathcal{A}_{\mathbf{r}^*}, \end{cases}$$

while $\mathbf{r} \in \mathcal{A}_{>}$ implies that $\mu(\lambda_2; \mathbf{r}) \equiv 0$. This last conclusion implies that (A.4) can be explicitly written as

$$g^n(\lambda_2) = \sum_{\mathbf{r} \in \Delta_{\omega_1+\omega_2, n} \cap \{\mathcal{A}_{\mathbf{r}^*} \cup \mathcal{A}_{\leq, \mathbf{0}}\}} \frac{n!}{r_1! \dots r_{\omega_1+\omega_2}!} \mu(\lambda_2; \mathbf{r}).$$

Now note that $\forall \mathbf{r} \in \mathcal{A}_{\leq, \mathbf{0}}$, $\sum_{k=1}^{\omega_1+\omega_2} r_k < \omega_1\omega_2$, while $\sum_{k=1}^{\omega_1+\omega_2} r_k^* = \omega_1\omega_2$. This shows that $\Delta_{\omega_1+\omega_2, n} \cap \mathcal{A}_{\leq, \mathbf{0}} = \emptyset$ for all $n \geq \omega_1\omega_2$, while $\Delta_{\omega_1+\omega_2, n} \cap \mathcal{A}_{\mathbf{r}^*} = \emptyset$ for all $n \neq \omega_1\omega_2$.

It is now trivial to see that $g^n(\lambda_2)|_{\lambda_2=\lambda_1} = 0 \ \forall 0 \leq n < \omega_1\omega_2$, $g^{\omega_1\omega_2}(\lambda_2)|_{\lambda_2=\lambda_1} = (\omega_1\omega_2)! \prod_{1 \leq k_1 < k_2 \leq \omega_1+\omega_2} (t_{k_2} - t_{k_1})$ and $g^n(\lambda_2) \equiv 0$ for all $n \geq \omega_1\omega_2 + 1$. This establishes the desired result. \square

The general version of the above result is the following.

LEMMA A.2. *For a fixed integer $d \geq 2$, let $\omega_1, \dots, \omega_d$ be arbitrary natural numbers. Let $S_0 = 0$ and for $k \in [d]$, $S_k = S_{k-1} + \omega_k$. Further suppose that $\lambda_1, \dots, \lambda_d$ and t_1, \dots, t_{S_d} are strictly positive real numbers satisfying $\lambda_{i_1} \neq \lambda_{i_2}$, if $i_1 \neq i_2$ and $t_{j_1} \neq t_{j_2}$, if $j_1 \neq j_2$. Then the square matrix*

$$(A.5) \quad W = \left(\begin{array}{ccc|ccc} \frac{1}{(\lambda_1+t_1)^{\omega_1}} & \cdots & \frac{1}{(\lambda_1+t_1)} & \vdots & \cdots & \frac{1}{(\lambda_d+t_1)^{\omega_d}} & \cdots & \frac{1}{(\lambda_d+t_1)} \\ \vdots & \ddots & \vdots & \vdots & \ddots & \vdots & \ddots & \vdots \\ \frac{1}{(\lambda_1+t_{S_d})^{\omega_1}} & \cdots & \frac{1}{(\lambda_1+t_{S_d})} & \vdots & \cdots & \frac{1}{(\lambda_d+t_{S_d})^{\omega_d}} & \cdots & \frac{1}{(\lambda_d+t_{S_d})} \end{array} \right)$$

is non-singular.

Proof. It suffices to show that

$$(A.6) \quad \det(W) = \left(\prod_{j=1}^{S_d} \prod_{i=1}^d \frac{\prod_{(i_1, i_2) \in \mathcal{I}_d} (\lambda_{i_2} - \lambda_{i_1})^{\omega_{i_1} \omega_{i_2}}}{(\lambda_i + t_j)^{\omega_i}} \right) \prod_{(j_1, j_2) \in \mathcal{J}} (t_{j_2} - t_{j_1}),$$

where $\mathcal{I}_d = \{(i_1, i_2) : i_1, i_2 \in [d], i_1 < i_2\}$ and $\mathcal{J} = \{(j_1, j_2) : j_1, j_2 \in [S_d], j_1 < j_2\}$.

As a first step, we do the row operations

$$\text{row}_j \leftarrow \text{row}_j \times \prod_{i=1}^d (\lambda_i + t_j)^{\omega_i}$$

for each $j \in [S_d]$ on W to get the matrix B . For $j, k \in [S_d]$, note that

$$(A.7) \quad (B)_{jk} = \begin{cases} \left\{ \prod_{i \in [d] \setminus \{1\}} (\lambda_i + t_j)^{\omega_i} \right\} (\lambda_1 + t_j)^{k - S_0 - 1}, & S_0 + 1 \leq k \leq S_1 \\ \vdots \\ \left\{ \prod_{i \in [d] \setminus \{d\}} (\lambda_i + t_j)^{\omega_i} \right\} (\lambda_d + t_j)^{k - S_{(d-1)} - 1}, & S_{(d-1)} + 1 \leq k \leq S_d. \end{cases}$$

To verify (A.6), it suffices to show that

$$(A.8) \quad \det(B) = \left(\prod_{(i_1, i_2) \in \mathcal{I}_d} (\lambda_{i_2} - \lambda_{i_1})^{\omega_{i_1} \omega_{i_2}} \right) \prod_{(j_1, j_2) \in \mathcal{J}} (t_{j_2} - t_{j_1}).$$

For a fixed d , let \mathcal{C}_d denote the collection of matrices of the form similar to B for all possible choices of $\omega_1, \dots, \omega_d, \lambda_1, \dots, \lambda_d, t_1, \dots, t_{S_d}$ satisfying the given conditions of the lemma. We will say \mathcal{C}_d is non-singular if the determinant of each matrix in this collection is given by (A.8) and hence is non-zero. Now let $\mathcal{N} := \{d \in \mathbb{Z}_{++} : \mathcal{C}_d \text{ is non-singular}\}$. In these notations, a claim equivalent to (A.8) is to show that if $d \geq 2$, then $d \in \mathcal{N}$. We prove this alternate claim using induction.

From Lemma A.1 it follows that $2 \in \mathcal{N}$. We treat this as our base case. Let the induction hypothesis be that $d - 1 \in \mathcal{N}$ for some $d \geq 3$. To check if $d \in \mathcal{N}$, we verify whether (A.8) holds true for the determinant of an arbitrary matrix in \mathcal{C}_d . For convenience, we reuse the symbol B to denote this arbitrary matrix. In relation to B , let $g(\lambda_d), g^n(\lambda_d), b_k(\lambda_d), b_k^{r_k}(\lambda_d), \mathbf{r} \equiv (r_1, \dots, r_{S_d}) \in \mathbb{Z}_+^{S_d}, \Delta_{S_d, n}$ and $\mu(\lambda_d; \mathbf{r})$ be defined as in the proof of Lemma A.1.

Our approach is similar in spirit to that used in verifying (A.3). That is, we treat λ_d as a variable and the other indeterminates as constants and, using the induction hypothesis, show that

- i. $g(\lambda_d)$ is a univariate polynomial of degree $S_{d-1}\omega_d$ and for each $m \in [d-1]$, λ_m is a root with multiplicity exactly $\omega_m\omega_d$, i.e., $g(\lambda_d) = h \times \prod_{i \in [d-1]} (\lambda_d - \lambda_i)^{\omega_i\omega_d}$.

$$\text{ii. } h = \left(\prod_{(i_1, i_2) \in \mathcal{I}_{(d-1)}} (\lambda_{i_2} - \lambda_{i_1})^{\omega_{i_1} \omega_{i_2}} \right) \times \left(\prod_{(j_1, j_2) \in \mathcal{J}} (t_{j_2} - t_{j_1}) \right).$$

To begin with, observe that

$$(A.9) \quad g^n(\lambda_d) = \sum_{\mathbf{r} \in \Delta_{S_d, n}} \frac{n!}{r_1! \dots r_{S_d}!} \mu(\lambda_d; \mathbf{r}).$$

Let us now fix an arbitrary $m \in [d-1]$. Taking hints from the observations made in Lemma A.1, we define $\mathbf{r}^* \in \mathbb{Z}_+^{S_d}$ by

$$r_k^* = \begin{cases} \omega_d, & S_{(m-1)} + 1 \leq k \leq S_m \\ 0, & \text{otherwise.} \end{cases}$$

We next partition $\mathbb{Z}_+^{S_d}$ into the three sets $\mathcal{A}_{\leq, \mathbf{0}}$, $\mathcal{A}_{\mathbf{r}^*}$, and $\mathcal{A}_{>}$ given respectively by

$$\mathcal{A}_{\leq, \mathbf{0}} = \{ \mathbf{r} : 0 \leq r_{S_{(m-1)}+1} < \omega_d; \quad \forall 2 \leq k_1 \leq \omega_m, \quad r_{S_{(m-1)}+k_1} \in \{0\} \cup [\omega_d] \\ \text{and } r_{k_2} = 0 \text{ if } S_{(d-1)} + 1 \leq k_2 \leq S_d \},$$

$\mathcal{A}_{\mathbf{r}^*} = \{\mathbf{r}^*\}$ and $\mathcal{A}_{>} = \mathbb{Z}_+^{S_d} \setminus \{\mathcal{A}_{\mathbf{r}^*} \cup \mathcal{A}_{\leq, \mathbf{0}}\}$. With respect to $\mathcal{A}_{\leq, \mathbf{0}}$ note that, when $1 \leq k \leq S_{(m-1)}$ or $S_m + 1 \leq k \leq S_{(d-1)}$, there is no restriction on r_k . This necessarily implies that for each $\mathbf{r} \in \mathcal{A}_{>}$, $\sum_k r_k > \omega_d \omega_m$. Hence these tuples do not appear in the expansion of $g^n(\lambda_d)$, as given in (A.9), whenever $n \leq \omega_m \omega_d$. So we ignore $\mathcal{A}_{>}$ for the time being and determine the value of $\mu(\lambda_d; \mathbf{r})$ for tuples lying in the other two sets using the definition of B given in (A.7).

1. $\mathbf{r} \in \mathcal{A}_{\leq, \mathbf{0}}$: We consider the following subcases.

(a) $r_{S_m} < \omega_d$: Here observe that when $\lambda_d = \lambda_m$,

$$b_{S_m}^{r_{S_m}}(\lambda_d)|_{\lambda_d=\lambda_m} = \begin{cases} b_{S_d}(\lambda_d)|_{\lambda_d=\lambda_m}, & r_{S_m} = 0 \\ \left(\prod_{j=0}^{r_{S_m}-1} (\omega_d - j) \right) \times b_{S_d-r_{S_m}}(\lambda_d)|_{\lambda_d=\lambda_m}, & r_{S_m} \in [\omega_d - 1]. \end{cases}$$

(b) There exists $k \in [\omega_m - 1]$ such that $r_{S_m} = r_{S_m-1} = \dots = r_{S_m-k+1} = \omega_d$ and $r_{S_m-k} < \omega_d$: For this subcase, observe that if $\omega_m \leq \omega_d$, or $\omega_m > \omega_d$ and $1 \leq k < \omega_d$, then

$$b_{S_m-k}^{r_{(S_m-k)}}(\lambda_d)|_{\lambda_d=\lambda_m} = \begin{cases} b_{S_d-k}(\lambda_d)|_{\lambda_d=\lambda_m}, & r_{(S_m-k)} = 0 \\ \left(\prod_{j=0}^{r_{(S_m-k)}-1} (\omega_d - j) \right) \times b_{S_d-k-r_{(S_m-k)}}(\lambda_d)|_{\lambda_d=\lambda_m}, & 1 \leq r_{(S_m-k)} < \omega_d - k \\ \frac{b_{S_m+\omega_d-k-r_{(S_m-k)}}^{\omega_d}(\lambda_d)|_{\lambda_d=\lambda_m}}{(\omega_d - r_{S_m-k})!}, & \omega_d - k \leq r_{S_m-k} < \omega_d \end{cases}$$

On the other hand, if $\omega_m > \omega_d$ and $\omega_d \leq k < \omega_m$, then

$$b_{S_m-k}^{r_{(S_m-k)}}(\lambda_d)|_{\lambda_d=\lambda_m} = \frac{b_{S_m+\omega_d-k-r_{(S_m-k)}}^{\omega_d}(\lambda_d)|_{\lambda_d=\lambda_m}}{(\omega_d - r_{S_m-k})!}, \quad 0 \leq r_{S_m-k} < \omega_d.$$

From this, it follows that for each $\mathbf{r} \in \mathcal{A}_{\leq, \mathbf{0}}$ there exists a pair of columns which are linearly dependent when $\lambda_d = \lambda_m$ and thus $\mu(\lambda_d; \mathbf{r})|_{\lambda_d=\lambda_m} = 0$.

2. $\mathbf{r} \in \mathcal{A}_{\mathbf{r}^*}$: Let V denote the matrix obtained after differentiating element-wise the individual columns of B up to orders as indicated by \mathbf{r}^* and substituting $\lambda_d = \lambda_m$. It is easy to see that for $j, k \in [S_d]$,

$$(V)_{jk} = \begin{cases} \left(\prod_{i \in [d] \setminus \{b, m\}} (\lambda_i + t_j)^{\omega_i} \right) \times (\lambda_b + t_j)^{k - S_{(b-1)} - 1} (\lambda_m + t_j)^{\omega_m + \omega_d}, & S_{(b-1)} < k \leq S_b \\ & \text{for some } b \in [d] \setminus \{m, d\}. \\ \omega_d! \left(\prod_{i \in [d] \setminus \{m, d\}} (\lambda_i + t_j)^{\omega_i} \right) \times (\lambda_m + t_j)^{k - S_{(m-1)} - 1}, & S_{(m-1)} < k \leq S_m \\ \left(\prod_{i \in [d] \setminus \{m, d\}} (\lambda_i + t_j)^{\omega_i} \right) \times (\lambda_m + t_j)^{\omega_m + k - S_{(d-1)} - 1}, & S_{(d-1)} < k \leq S_d. \end{cases}$$

Now observe that if we define the matrix \tilde{V} from V as

$$(\tilde{V})_{jk} = \begin{cases} (V)_{jk}, & 1 \leq k \leq S_{(m-1)}, \\ \frac{1}{\omega_d!} (V)_{jk}, & S_{(m-1)} + 1 \leq k \leq S_m \\ (V)_{j(k - S_m + S_{(d-1)})}, & S_m + 1 \leq k \leq S_m + \omega_d \\ (V)_{j(k - \omega_d)}, & S_m + \omega_d + 1 \leq k \leq S_d, \end{cases}$$

then $\tilde{V} \in \mathcal{C}_{d-1}$. From the induction hypothesis, it follows that

$$\det \tilde{V} = (-1)^{S_{(d-1)} - S_{(m-1)}} \left(\prod_{j=1}^{d-1} (\lambda_m - \lambda_j)^{\omega_j \omega_d} \right) \left(\prod_{(j_1, j_2) \in \mathcal{J}} (t_{j_2} - t_{j_1}) \right) \times \left(\prod_{(i_1, i_2) \in \mathcal{I}_{(d-1)}} (\lambda_{i_2} - \lambda_{i_1})^{\omega_{i_1} \omega_{i_2}} \right).$$

Consequently,

$$\begin{aligned} \mu(\lambda_d; \mathbf{r}^*)|_{\lambda_d = \lambda_m} &= \det(V) \\ &= \left(\prod_{j=1}^{d-1} (\lambda_m - \lambda_j)^{\omega_j \omega_d} \right) \left(\prod_{(j_1, j_2) \in \mathcal{J}} (t_{j_2} - t_{j_1}) \right) \times \\ &\quad \left(\prod_{(i_1, i_2) \in \mathcal{I}_{(d-1)}} (\lambda_{i_2} - \lambda_{i_1})^{\omega_{i_1} \omega_{i_2}} \right). \end{aligned}$$

Now for each $n < \omega_m \omega_d$, note that $\Delta_{S_d, n} \subset \mathcal{A}_{\leq, 0}$ while for $n = \omega_m \omega_d$, $\Delta_{S_d, n} \subset \mathcal{A}_{\leq, 0} \cup \mathcal{A}_{\mathbf{r}^*}$. From the above observations and (A.9), it then follows that, for each

$0 \leq n < \omega_m \omega_d$, $g^n(\lambda_d)|_{\lambda_d=\lambda_m} = 0$ while

$$(A.10) \quad g^{\omega_m \omega_d}(\lambda_d)|_{\lambda_d=\lambda_m} = (\omega_m \omega_d)! \left(\prod_{(i_1, i_2) \in \mathcal{I}_{(d-1)}} (\lambda_{i_2} - \lambda_{i_1})^{\omega_{i_1} \omega_{i_2}} \right) \times \left(\prod_{(j_1, j_2) \in \mathcal{J}} (t_{j_2} - t_{j_1}) \right) \left(\prod_{j=1}^{d-1} (\lambda_m - \lambda_j)^{\omega_j \omega_d} \right).$$

Clearly, $g^{\omega_m \omega_d}(\lambda_d)|_{\lambda_d=\lambda_m} \neq 0$. Thus, λ_m is a root of $g(\lambda_d)$ of multiplicity $\omega_m \omega_d$.

Since m was arbitrary to start with, it follows that for every $i \in [d-1]$, $(\lambda_d - \lambda_i)^{\omega_i \omega_d}$ is a factor of $g(\lambda_d)$. This suggests that original structure of $g(\lambda_d)$ must have been of the form $g(\lambda_d) = h(\lambda_d) \prod_{i \in [d-1]} (\lambda_d - \lambda_i)^{\omega_i \omega_d}$ for some univariate polynomial $h(\lambda_d)$. The fact that the determined multiplicities of the roots of $g(\lambda_d)$ are exact ensures that $h(\lambda_i) \neq 0 \forall i \in [d-1]$. Thus h is not an identically zero polynomial.

It remains to show that h is a constant and in particular

$$(A.11) \quad h = \left(\prod_{(i_1, i_2) \in \mathcal{I}_{(d-1)}} (\lambda_{i_2} - \lambda_{i_1})^{\omega_{i_1} \omega_{i_2}} \right) \times \left(\prod_{(j_1, j_2) \in \mathcal{J}} (t_{j_2} - t_{j_1}) \right).$$

Towards this observe that if $n \geq S_{(d-1)}\omega_d + 1$, then for every tuple $\mathbf{r} \in \Delta_{S_d, n}$ either one of $r_{S_{(d-1)}+1}, \dots, r_{S_d}$ is strictly greater than zero or one amongst $r_1, \dots, r_{S_{(d-1)}}$ is strictly bigger than ω_d . In the former situation, arguing as in observations (1) and (2) from Lemma (A.1), it is easy to see that $\mu(\lambda_d; \mathbf{r}) \equiv 0$. For the latter situation, let us suppose that for some arbitrary $1 \leq k_0 \leq S_{(d-1)}$, $r_{k_0} > \omega_d$. The fact that every element $(B)_{jk_0}$ of column k_0 is a polynomial in λ_d of degree ω_d immediately shows that $b_{k_0}^{r_{k_0}}(\lambda_d) \equiv 0$ and hence $\mu(\lambda_d; \mathbf{r}) \equiv 0$. Consequently, it follows that $g^n(\lambda_d) \equiv 0 \forall n > S_{(d-1)}\omega_d$. That is, the degree of $g(\lambda_d)$ is exactly $S_{(d-1)}\omega_d$. In other words,

$$(A.12) \quad g(\lambda_d) = h \times \prod_{i \in [d-1]} (\lambda_d - \lambda_i)^{\omega_i \omega_d}$$

for some constant $h \neq 0$. Differentiating (A.12) up to order $\omega_m \omega_d$ for some arbitrary $m \in [d-1]$ and comparing with (A.10), (A.11) immediately follows as desired. \square

We now finally prove Lemma 4.2 by showing the following general result.

THEOREM A.3. *For a fixed integer $d \geq 2$, let $\omega_1, \dots, \omega_d$ be arbitrary natural numbers. Let $S_0 = 0$ and for $k \in [d]$, $S_k = S_{k-1} + \omega_k$. Further suppose that $\lambda_1, \dots, \lambda_d$ and t_1, \dots, t_{S_d} are strictly positive real numbers satisfying $\lambda_{i_1} \neq \lambda_{i_2}$, if $i_1 \neq i_2$ and $t_{j_1} \neq t_{j_2}$, if $j_1 \neq j_2$. Then the square matrix*

$$(A.13) \quad M := \left(\begin{array}{ccc|ccc} \frac{t_1}{(\lambda_1 + t_1)} & \cdots & \frac{t_1^{\omega_1}}{(\lambda_1 + t_1)^{\omega_1}} & \cdots & \frac{t_1}{(\lambda_d + t_1)} & \cdots & \frac{t_1^{\omega_d}}{(\lambda_d + t_1)^{\omega_d}} \\ \vdots & \ddots & \vdots & \cdots & \vdots & \ddots & \vdots \\ \frac{t_{S_d}}{(\lambda_1 + t_{S_d})} & \cdots & \frac{t_{S_d}^{\omega_1}}{(\lambda_1 + t_{S_d})^{\omega_1}} & \cdots & \frac{t_{S_d}}{(\lambda_d + t_{S_d})} & \cdots & \frac{t_{S_d}^{\omega_d}}{(\lambda_d + t_{S_d})^{\omega_d}} \end{array} \right),$$

is non-singular.

Proof. For $k \in [S_d]$, let $b_k := \min\{i \in [d] : S_i \geq k\}$. Now using the expansion $\lambda^{n-1} = \sum_{q=0}^{n-1} (-1)^q \binom{n-1}{q} t^q (\lambda + t)^{n-1-q}$, for $n \in \mathbb{Z}_{++}$, observe that

$$(A.14) \quad \sum_{q=0}^{n-1} (-1)^q \binom{n-1}{q} \frac{t^{q+1}}{(\lambda + t)^{q+1}} = \frac{\lambda^{n-1} t}{(\lambda + t)^n}.$$

Using this note that, for any $j, k \in [S_d]$,

$$(A.15) \quad \sum_{q=0}^{k-S_{(b_k-1)}-1} (-1)^q \binom{k-S_{(b_k-1)}-1}{q} (M)_{j(S_{(b_k-1)}+1+q)} = \frac{\lambda_{b_k}^{k-S_{(b_k-1)}-1} t_j}{(\lambda_{b_k} + t_j)^{k-S_{(b_k-1)}}}.$$

From this it follows that if, for each $i \in [d]$, we perform the column operations

$$\text{col}_k \leftarrow \sum_{q=0}^{k-S_{(i-1)}-1} (-1)^q \binom{k-S_{(i-1)}-1}{q} \text{col}_{S_{(i-1)}+1+q}$$

in the order $k = S_i, S_i - 1, \dots, S_{(i-1)} + 1$, then we end up with matrix

$$U = \left(\begin{array}{ccc|ccc} \frac{t_1}{(\lambda_1+t_1)} & \cdots & \frac{\lambda_1^{\omega_1-1} t_1}{(\lambda_1+t_1)^{\omega_1}} & \cdots & \frac{t_1}{(\lambda_d+t_1)} & \cdots & \frac{\lambda_d^{\omega_d-1} t_1}{(\lambda_d+t_1)^{\omega_d}} \\ \vdots & \ddots & \vdots & \cdots & \vdots & \ddots & \vdots \\ \frac{t_{S_d}}{(\lambda_1+t_{S_d})} & \cdots & \frac{\lambda_1^{\omega_1-1} t_{S_d}}{(\lambda_1+t_{S_d})^{\omega_1}} & \cdots & \frac{t_{S_d}}{(\lambda_d+t_{S_d})} & \cdots & \frac{\lambda_d^{\omega_d-1} t_{S_d}}{(\lambda_d+t_{S_d})^{\omega_d}} \end{array} \right).$$

Since only reversible column operations are used to obtain U from M , it suffices to show that U is non-singular. But this is true from Lemma A.2. The desired result thus follows. \square

Appendix B. Coefficient expansion.

LEMMA B.1. *For any $\omega \in \mathbb{Z}_+^d$, if $\mathcal{D}(\omega) \neq \emptyset$, then*

$$(B.1) \quad \Lambda^\omega = \sum_{k \in \mathcal{D}(\omega)} \sum_{q=1}^{\omega_k} \gamma_{kq}(\omega) \Lambda_k^q(t)$$

for all t . Further, this expansion is unique.

Proof. The uniqueness of the expansion is a simple consequence of Lemma A.13. We prove (B.1) using induction on $\deg(\omega) = \sum_{k=1}^d \omega_k$.

For $\deg(\omega) = 2$, our basis step, we verify (B.1) by considering the following exhaustive cases.

1. there exists a unique $k \in [d]$ such that $\omega_k = 2$: Here, with $\mathcal{D}(\omega) = \{k\}$, observe that $\bar{\Delta}_{k1}(\omega) = \emptyset$ and $\bar{\Delta}_{k2}(\omega) = \{\mathbf{0} \in \mathbb{Z}_+^d\}$. Consequently, $\gamma_{11}(\omega) = 0$ and $\gamma_{12}(\omega) = 1$. Equation (B.1) thus trivially holds.
2. there exists unique $k_1, k_2 \in [d]$ such that $\omega_{k_1} = \omega_{k_2} = 1$: Here observe that $\mathcal{D}(\omega) = \{k_1, k_2\}$, $\bar{\Delta}_{k_11}(\omega) = \bar{\Delta}_{k_21}(\omega) = \{\mathbf{0} \in \mathbb{Z}_+^d\}$, $\gamma_{k_11}(\omega) = \beta_{k_1 k_2}$ and $\gamma_{k_21}(\omega) = \beta_{k_2 k_1}$. Equation (B.1) clearly holds true for this case since $\forall t$

$$\begin{aligned} \Lambda^\omega &= \Lambda_{k_1}(t) \Lambda_{k_2}(t) \\ &= \beta_{k_1 k_2} \Lambda_{k_1}(t) + \beta_{k_2 k_1} \Lambda_{k_2}(t) \\ &= \gamma_{k_11}(\omega) \Lambda_{k_1}(t) + \gamma_{k_21}(\omega) \Lambda_{k_2}(t). \end{aligned}$$

Now for some fixed $n \geq 2$, let the strong induction hypothesis be that (B.1) holds for all ω such that $\deg(\omega) \leq n$. To verify (B.1) when $\deg(\omega) = n+1$, we consider the following exhaustive cases.

1. there exists a unique $k \in [d]$ such that $\omega_k = n+1$: Here $\mathcal{D}(\omega) = \{k\}$,

$$\bar{\Delta}_{kq}(\omega) = \begin{cases} \emptyset, & q < n+1 \\ \{\mathbf{0} \in \mathbb{Z}_+^d\}, & q = n+1 \end{cases}$$

and consequently

$$\gamma_{kq}(\omega) = \begin{cases} 0, & q < n+1 \\ 1, & q = n+1. \end{cases}$$

Using these note that (B.1) again holds trivially.

2. there exist unique $k_1, k_2 \in [d]$ such that $\omega_{k_1}, \omega_{k_2} > 0$ and $\omega_{k_1} + \omega_{k_2} = n+1$: The fact that $n \geq 2$ immediately ensures that at least one of $\omega_{k_1}, \omega_{k_2}$ is strictly bigger than unity. Without loss of generality, let us assume that $k_1 = 1, k_2 = 2$ and $\omega_1 > 1$. Observe then that $\mathcal{D}(\omega) = \{1, 2\}$. Also, for each $q \in [\omega_1]$, we have $\bar{\Delta}_{1q}(\omega) = \{(0, \omega_1 - q, 0, \dots, 0) \in \mathbb{Z}_+^d\}$, a singleton set. Hence $\gamma_{1q}(\omega) = \beta_{12}^{\omega_2} \binom{\omega_2 + \omega_1 - q - 1}{\omega_2 - 1} \beta_{21}^{\omega_1 - q}$. Similarly, $\forall q \in [\omega_2]$, $\gamma_{2q}(\omega) = \beta_{21}^{\omega_1} \binom{\omega_1 + \omega_2 - q - 1}{\omega_1 - 1} \beta_{12}^{\omega_2 - q}$. Observe next that if $\omega'_1 := \omega_1 - 1$, a positive number, then Λ^ω can be written as $\Lambda_1(t) \left\{ \Lambda^{\omega'} \right\}$, where $\omega' = (\omega'_1, \omega_2, 0, \dots, 0)$. Clearly, $\deg(\omega') = n$, $\gamma_{1q}(\omega') = \beta_{12}^{\omega_2} \binom{\omega_2 + \omega'_1 - q - 1}{\omega_2 - 1} \beta_{21}^{\omega'_1 - q}$, $\forall q \in [\omega'_1]$, and $\gamma_{2q}(\omega') = \beta_{21}^{\omega'_1} \binom{\omega'_1 + \omega_2 - q - 1}{\omega'_1 - 1} \beta_{12}^{\omega_2 - q}$, $\forall q \in [\omega_2]$. By induction hypothesis, it then follows that

$$\begin{aligned} \Lambda^\omega &= \sum_{q=1}^{\omega'_1} \gamma_{1q}(\omega') \Lambda_1^{q+1}(t) + \sum_{q=1}^{\omega_2} \gamma_{2q}(\omega') \Lambda_1(t) \Lambda_2^q(t) \\ (B.2) \quad &:= \text{term}_1 + \text{term}_2. \end{aligned}$$

Since $\omega'_1 = \omega_1 - 1$, it follows that

$$\begin{aligned} \text{term}_1 &= \sum_{q=1}^{\omega'_1} \beta_{12}^{\omega_2} \binom{\omega_2 + \omega'_1 - q - 1}{\omega_2 - 1} \beta_{21}^{\omega'_1 - q} \Lambda_1^{q+1}(t) \\ &= \sum_{q=2}^{\omega_1} \beta_{12}^{\omega_2} \binom{\omega_2 + \omega_1 - q - 1}{\omega_2 - 1} \beta_{21}^{\omega_1 - q} \Lambda_1^q(t) \\ (B.3) \quad &= \sum_{q=2}^{\omega_1} \gamma_{1q}(\omega) \Lambda_1^q(t). \end{aligned}$$

For term_2 , we consider two subcases.

- (a) $\omega_2 = 1$ or equivalently $\omega_1 = n$: Here, $\text{term}_2 = \beta_{21}^{n-1} \Lambda_1(t) \Lambda_2(t)$. By additionally using the base case, it follows that

$$\begin{aligned} \text{term}_2 &= \beta_{21}^{n-1} \beta_{12} \Lambda_1(t) + \beta_{21}^n \Lambda_2(t) \\ (B.4) \quad &= \gamma_{11}(\omega) \Lambda_1(t) + \gamma_{21}(\omega) \Lambda_2(t) \end{aligned}$$

Substituting (B.3) and (B.4) in (B.2), it is clear that (B.1) holds.

- (b) $\omega_2 > 1$: We again apply the induction hypothesis individually to the term $\Lambda_1(t) \Lambda_2^q(t)$ for each $1 \leq q \leq \omega_2$. Note that none of these expansions yield scalar multiples of $\Lambda_1^q(t)$ for any $2 \leq q \leq \omega_1$ and, thus, their associated weights obtained in (B.3) remain unchanged in the overall decomposition of Λ^ω . The term $\Lambda_1(t)$ is, however, definitely present in the final decomposition of term_2 . For each $q \in [\omega_2]$, observe that the constant associated with $\Lambda_1(t)$ in the expansion of $\Lambda_1(t) \Lambda_2^q(t)$ is $\gamma_{11}(1, q, 0, \dots, 0) = \beta_{12}^q$. From the definition of term_2 ,

clearly, the constant associated to $\Lambda_1(t)$ in the final decomposition of Λ^ω is

$$\begin{aligned} \sum_{q=1}^{\omega_2} \gamma_{2q}(\omega') \beta_{12}^q &= \beta_{21}^{\omega_1'} \beta_{12}^{\omega_2} \sum_{q=1}^{\omega_2} \binom{\omega_1' + \omega_2 - q - 1}{\omega_1' - 1} \\ &= \beta_{21}^{\omega_1'} \beta_{12}^{\omega_2} \binom{\omega_1' + \omega_2 - 1}{\omega_1' - 1}, \\ &= \beta_{21}^{\omega_1 - 1} \beta_{12}^{\omega_2} \binom{\omega_2 + \omega_1 - 1 - 1}{\omega_2 - 1} \\ &= \gamma_{11}(\omega). \end{aligned}$$

Note that the second equality above is a consequence of the well known binomial identity $\sum_{r=j}^n \binom{r}{j} = \binom{n+1}{j+1}$. Next observe that the decomposition of $\Lambda_1(t) \Lambda_2^q(t)$ by the induction hypothesis also results in a linear combination of $\Lambda_2(t), \Lambda_2^2(t), \dots, \Lambda_2^q(t)$. This implies that

$$(B.5) \quad \text{term}_2 = \gamma_{11}(\omega) \Lambda_1(t) + \sum_{q=1}^{\omega_2} v_q \Lambda_2^q(t)$$

for some t -independent constant $\{v_q\}$. By substituting (B.3) and (B.5) in (B.2), it follows that

$$(B.6) \quad \Lambda^\omega = \sum_{q=1}^{\omega_1} \gamma_{1q}(\omega) \Lambda_1^q(t) + \sum_{q=1}^{\omega_2} v_q \Lambda_2^q(t).$$

Now recall that $\omega_2 > 1$ in the subcase under consideration. Expressing Λ^ω as $\Lambda_2(t) \{ \Lambda_1^{\omega_1}(t) \Lambda_2^{\omega_2-1}(t) \}$ and interchanging the roles of $\Lambda_1(t)$ and $\Lambda_2(t)$ in the above argument, it then follows that

$$(B.7) \quad \Lambda^\omega = \sum_{q=1}^{\omega_1} u_q \Lambda_1^q(t) + \sum_{q=1}^{\omega_2} \gamma_{2q}(\omega) \Lambda_2^q(t)$$

for some t -independent constants $\{u_q\}$. Comparing (B.6) and (B.7), we get

$$(B.8) \quad \sum_{q=1}^{\omega_1} \{ \gamma_{1q}(\omega) - u_q \} \Lambda_1^q(t) + \sum_{q=1}^{\omega_2} \{ v_q - \gamma_{2q}(\omega) \} \Lambda_2^q(t) = 0$$

for all t . A simple application of Lemma A.13 then shows that $u_q = \gamma_{1q}(\omega)$, $\forall q \in [\omega_1]$, and $v_q = \gamma_{2q}(\omega)$, $\forall q \in [\omega_2]$. By substituting these values in either (B.6) or (B.7), it follows that (B.1) holds for this case also.

3. Cardinality of the set $\mathcal{D}(\omega) = \{k \in [d] : \omega_k > 0\}$ is atleast 3 : Without loss of generality, let us suppose that $\omega_1, \omega_2 > 0$. Our approach here is to express Λ^ω as $\Lambda_1^{\omega_1}(t) \{ \Lambda_2^{\omega_2}(t) \dots \Lambda_d^{\omega_d}(t) \}$ and expand the term within the braces using the induction hypothesis. With $\omega'' = (0, \omega_2, \dots, \omega_d)$, it follows that

$$\begin{aligned} \Lambda^\omega &= \Lambda_1^{\omega_1}(t) \left\{ \sum_{k \in \mathcal{D}(\omega) \setminus \{1\}} \sum_{q=1}^{\omega_k} \gamma_{kq}(\omega'') \Lambda_k^q(t) \right\} \\ &= \sum_{k \in \mathcal{D}(\omega) \setminus \{1\}} \sum_{q=1}^{\omega_k} \gamma_{kq}(\omega'') \Lambda_1^{\omega_1}(t) \Lambda_k^q(t) \\ &:= \sum_{k \in \mathcal{D}(\omega) \setminus \{1\}} \text{term}_k. \end{aligned}$$

A neat observation, by virtue of our induction hypothesis, is that the eventual constant associated with $\Lambda_k^q(t)$, $k \in \mathcal{D}(\omega) \setminus \{1\}$, $q \in [\omega_k]$, depends solely on term_k . Infact the only terms that contribute a scalar multiple of $\Lambda_k^q(t)$ are $\Lambda_1^{\omega_1}(t)\Lambda_k^q(t)$, $\Lambda_1^{\omega_1}(t)\Lambda_k^{q+1}(t), \dots, \Lambda_1^{\omega_1}(t)\Lambda_k^{\omega_k}(t)$. With this in mind, we focus on term_2 and evaluate the eventual constant, say α , associated with $\Lambda_2^q(t)$ for some arbitrary $q \in [\omega_2]$. Firstly observe that the scalar multiple of $\Lambda_2^q(t)$ in the expansion of $\Lambda_1^{\omega_1}(t)\Lambda_2^{q+j}(t)$, when $0 \leq j \leq \omega_2 - q$, is $\gamma_{2q}(\omega_1, q+j, 0, \dots, 0)$. This implies that

$$\begin{aligned} \alpha &= \sum_{j=0}^{\omega_2-q} \gamma_{2(q+j)}(\omega'') \gamma_{2q}(\omega_1, q+j, 0, \dots, 0) \\ &:= \sum_{j=0}^{\omega_2-q} c_j. \end{aligned}$$

In relation to each c_j , observe that $\bar{\Delta}_{2q}(\omega_1, q+j, 0, \dots, 0) = \{(j, 0, \dots, 0)\}$, a singleton set, and for any $\mathbf{s}' \equiv (s'_1, \dots, s'_d) \in \bar{\Delta}_{2(q+j)}(\omega'')$, $s'_1 = 0$ and $\sum_{r=2}^d s'_r = \omega_2 - q - j$. This implies that $\mathbf{s}' \in \bar{\Delta}_{2(q+j)}(\omega'')$ if and only if the tuple $(j, s'_2, s'_3, \dots, s'_d) \in \bar{\Delta}_{2q}(\omega) \cap \mathcal{B}_j$, where $\mathcal{B}_j := \{\mathbf{s} \in \mathbb{Z}_+^d : s_1 = j\}$. Thus,

$$c_j = \left(\prod_{r \in \mathcal{D}(\omega)} \beta_{2r}^{\omega_r} \right) \sum_{\mathbf{s} \in \bar{\Delta}_{2q}(\omega) \cap \mathcal{B}_j} \prod_{r \in \mathcal{D}(\omega)} \binom{\omega_r + s_r - 1}{\omega_r - 1} \beta_{r2}^{s_r}.$$

Consequently, it follows that $\alpha = \gamma_{2q}(\omega)$. By replicating this argument for every $k \in \mathcal{D} \setminus \{1\}$ and every $d \in [\omega_k]$, observe that the weight associated with $\Lambda_k^q(t)$ in the final expansion is $\gamma_{kq}(\omega)$. This, combined with the fact that the decomposition of $\Lambda_1^{\omega_1}(t)\Lambda_k^q(t)$ also results in a linear combination of $\Lambda_1(t), \dots, \Lambda_1^{\omega_1}(t)$, shows that

$$\Lambda^\omega = \sum_{q=1}^{\omega_1} u_q \Lambda_1^q(t) + \sum_{k \in \mathcal{D}(\omega) \setminus \{1\}} \sum_{q=1}^{\omega_k} \gamma_{kq}(\omega) \Lambda_k^q(t)$$

for some t -independent constants u_1, \dots, u_{ω_1} . But observe that Λ^ω can also be expressed as $\Lambda_2^{\omega_2}(t) \{\Lambda_1^{\omega_1}(t) \Lambda_3^{\omega_3}(t) \dots \Lambda_d^{\omega_d}(t)\}$. By repeating the above arguments with an interchange of the roles of $\Lambda_1(t)$ and $\Lambda_2(t)$, we, thus, also get

$$\Lambda^\omega = \sum_{q=1}^{\omega_2} v_q \Lambda_2^q(t) + \sum_{k \in \mathcal{D}(\omega) \setminus \{2\}} \sum_{q=1}^{\omega_k} \gamma_{kq}(\omega) \Lambda_k^q(t)$$

for some real constants v_1, \dots, v_{ω_2} . Arguing as in case (2b) above, it is easy to see that $u_q = \gamma_{1q}(\omega)$, $\forall q \in [\omega_1]$ and $v_q = \gamma_{2q}(\omega)$, $\forall q \in [\omega_2]$.

This completes the induction argument and, hence, proves the desired result. \square

Appendix C. Regularity of EPS. For discussions pertaining to this section, we suppose that the domain and the co-domain of the map \mathcal{E} , used to define the EPS of (4.13), is $\mathbb{C}^{d \cdot N_i}$. We now prove Lemma 4.4 through the following series of results.

LEMMA C.1. *Suppose a_1, \dots, a_{N_i} are N_i distinct complex numbers. Let $\mathbf{a} \equiv (\mathbf{a}_1, \dots, \mathbf{a}_{N_i})$, where $\mathbf{a}_j \equiv (a_j, \dots, a_j) \in \mathbb{C}^d$. Then*

$$(C.1) \quad \det(\dot{\mathcal{E}}(\mathbf{x}))|_{\mathbf{x}=\mathbf{a}} = \prod_{1 \leq j_1 < j_2 \leq N_i} (a_{j_1} - a_{j_2})^d \neq 0.$$

Proof. Pick an arbitrary $\Omega \in \Delta_{d+1, N_i}$, $j_2 \in [N_i]$ and $k_2 \in [d]$. We first show that

$$(C.2) \quad \left. \frac{\partial g(\mathbf{x}; \Omega)}{\partial x_{j_2 k_2}} \right|_{\mathbf{x}=\mathbf{a}} = \begin{cases} 0, & \omega_{k_2} = 0 \\ \nu(\Omega, k_2) e_{\eta(\omega)}(a_1, \dots, a_{j_2-1}, a_{j_2+1}, \dots, a_{N_i}), & \text{otherwise,} \end{cases}$$

where $\nu(\Omega; k_2)$ is a number, independent of choice of the vector \mathbf{a} and $j_2 \in [N_i]$, $\eta(\omega) = \left(\sum_{k=1}^d \omega_k \right) - 1 \in \mathbb{Z}_+$ and $e_l(\cdot)$, for $0 \leq l \leq N_i - 1$, is the l^{th} elementary symmetric polynomial in $N_i - 1$ variables.

Suppose that $\omega_{k_2} = 0$. Then clearly, $\forall \mathbf{b} \in \mathcal{B}_\Omega$, $\theta_{k_2}(\mathbf{b}) = 0$. This implies that $g(\mathbf{x}; \Omega)$ is a constant with respect to $x_{j_2 k_2}$. Thus, $\left. \frac{\partial g(\mathbf{x}; \Omega)}{\partial x_{j_2 k_2}} \right|_{\mathbf{x}=\mathbf{a}} = 0$ as desired.

Now suppose that $\omega_{k_2} \in [N_i]$. Clearly,

$$g(\mathbf{x}; \Omega) = x_{j_2 k_2} \sum_{\mathbf{b} \in \mathcal{B}_\Omega, b_{j_2} = k_2} \left(\prod_{j=1, j \neq j_2, b_j \neq d+1}^{N_i} x_{j b_j} \right) + \sum_{\mathbf{b} \in \mathcal{B}_\Omega, b_{j_2} \neq k_2} \left(\prod_{j=1, b_j \neq d+1}^{N_i} x_{j b_j} \right).$$

Hence,

$$(C.3) \quad \left. \frac{\partial g(\mathbf{x}; \Omega)}{\partial x_{j_2 k_2}} \right|_{\mathbf{x}=\mathbf{a}} = \sum_{\mathbf{b} \in \mathcal{B}_\Omega, b_{j_2} = k_2} \left(\prod_{j=1, j \neq j_2, b_j \neq d+1}^{N_i} a_j \right).$$

Now for any $\mathcal{S} \subseteq [N_i]$, such that $|\mathcal{S}| = \omega_{d+1}$, define the set $\mathcal{B}_\Omega(j_2, k_2, \mathcal{S}) \subseteq \mathcal{B}_\Omega$ by

$$\mathcal{B}_\Omega(j_2, k_2, \mathcal{S}) = \{ \mathbf{b} \in [d+1]^{N_i} : b_{j_2} = k_2; \forall j \in \mathcal{S}, b_j = d+1; \Theta(\mathbf{b}) = \Omega \}.$$

Its cardinality, given by

$$(C.4) \quad |\mathcal{B}_\Omega(j_2, k_2, \mathcal{S})| = \binom{N_i - 1 - \omega_{d+1}}{\omega_1, \dots, \omega_{(k_2-1)}, \omega_{k_2}-1, \omega_{(k_2+1)}, \dots, \omega_d, 0},$$

clearly depends on $|\mathcal{S}|$, but not on \mathcal{S} itself. Also, it is independent of choice of the vector \mathbf{a} and $j_2 \in [N_i]$. By defining

$$(C.5) \quad \nu(\Omega; k_2) = |\mathcal{B}_\Omega(j_2, k_2, \mathcal{S})|$$

and using the fact that $\{ \mathbf{b} \in \mathcal{B}_\Omega : b_{j_2} = k_2 \} = \bigcup_{\mathcal{S} \subseteq [N_i], |\mathcal{S}| = \omega_{d+1}} \mathcal{B}_\Omega(j_2, k_2, \mathcal{S})$, a disjoint union, observe from (C.3) that

$$\begin{aligned} \left. \frac{\partial g(\mathbf{x}; \Omega)}{\partial x_{j_2 k_2}} \right|_{\mathbf{x}=\mathbf{a}} &= \sum_{\mathcal{S} \subseteq [N_i], |\mathcal{S}| = \omega_{d+1}} \nu(\Omega; k_2) \left(\prod_{j=1, j \notin \mathcal{S} \cup \{j_2\}}^{N_i} a_j \right) \\ &= \nu(\Omega; k_2) e_{\eta(\omega)}(a_1, \dots, a_{(j_2-1)}, a_{(j_2+1)}, \dots, a_{N_i}), \end{aligned}$$

as desired. This now completes the verification of (C.2).

Now fix an arbitrary $k_1, k_2 \in [d]$ and $q_1, j_2 \in [N_i]$. As of consequence of (C.2), observe that

$$(C.6) \quad \left. \frac{\partial h_{k_1 q_1}(\mathbf{x})}{\partial x_{j_2 k_2}} \right|_{\mathbf{x}=\mathbf{a}} = \sum_{\substack{\Omega \in \Delta_{d+1, N_i}, \\ \omega_{k_1} \geq q_1, \\ \omega_{k_2} > 0}} \left[\frac{\gamma_{k_1 q_1}(\omega) \nu(\Omega; k_2)}{\lambda_{d+1}^{N_i - q_1 - \omega_{d+1}}} \times e_{\eta(\omega)}(a_1, \dots, a_{j_2-1}, a_{j_2+1}, \dots, a_{N_i}) \right].$$

By grouping the symmetric polynomials in the above equation that have the same degree, i.e., identical values of $\eta(\omega)$, we get

$$(C.7) \quad \left. \frac{\partial h_{k_1 q_1}(\mathbf{x})}{\partial x_{j_2 k_2}} \right|_{\mathbf{x}=\mathbf{a}} = \sum_{l=q_1-1}^{N_i-1} c_l(k_1, q_1, k_2) e_l(a_1, \dots, a_{j_2-1}, a_{j_2+1}, \dots, a_{N_i}),$$

where

$$c_l(k_1, q_1, k_2) := \sum_{\substack{\boldsymbol{\Omega} \in \Delta_{d+1, N_i}, \\ \omega_{k_1} \geq q_1, \omega_{k_2} > 0, \eta(\boldsymbol{\omega})=l}} \left(\frac{\gamma_{k_1 q_1}(\boldsymbol{\Omega})}{\lambda_{d+1}^{N_i - q_1 - \omega_{d+1}}} \right) \nu(\boldsymbol{\Omega}; k_2).$$

Using (4.6), (C.4) and (C.5), note in particular that

$$(C.8) \quad c_{q_1-1}(k_1, q_1, k_2) = \begin{cases} 1, & k_1 = k_2 \\ 0, & \text{otherwise.} \end{cases}$$

For every $k_1, k_2 \in [N_i]$, with

$$(C.9) \quad \Xi_{k_1, k_2} := \begin{pmatrix} \frac{\partial h_{k_1 1}(\mathbf{x})}{\partial x_{1 k_2}} & \dots & \frac{\partial h_{k_1 1}(\mathbf{x})}{\partial x_{N_i k_2}} \\ \vdots & \ddots & \vdots \\ \frac{\partial h_{k_1 N_i}(\mathbf{x})}{\partial x_{1 k_2}} & \dots & \frac{\partial h_{k_1 N_i}(\mathbf{x})}{\partial x_{N_i k_2}} \end{pmatrix},$$

consider the matrix

$$(C.10) \quad \Xi := \dot{\mathcal{E}}(\mathbf{x}) = \begin{pmatrix} \Xi_{1,1} & \dots & \Xi_{1,d} \\ \vdots & \ddots & \vdots \\ \Xi_{d,1} & \dots & \Xi_{d,d} \end{pmatrix}.$$

Using elementary row operations, our strategy is show that

$$(C.11) \quad \det(\dot{\mathcal{E}}(\mathbf{x})) = \det \begin{pmatrix} \Gamma & \mathbf{0} & \dots & \mathbf{0} \\ \mathbf{0} & \Gamma & \dots & \mathbf{0} \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{0} & \mathbf{0} & \dots & \Gamma \end{pmatrix} = (\det(\Gamma))^d,$$

where

$$(C.12) \quad \Gamma = \begin{pmatrix} e_0(a_2, \dots, a_{N_i}) & \dots & e_0(a_1, \dots, a_{N_i-1}) \\ \vdots & \ddots & \vdots \\ e_{N_i-1}(a_2, \dots, a_{N_i}) & \dots & e_{N_i-1}(a_1, \dots, a_{N_i-1}) \end{pmatrix}.$$

This is clearly sufficient to verify (C.1) since $\det(\Gamma) = \prod_{1 \leq j_1 < j_2 \leq N_i} (a_{j_1} - a_{j_2})$ is a well known result.

For notational convenience, the matrix obtained after applying elementary operations on Ξ will again be referred to by Ξ . The notation $\text{row}(r, k_1)$ will mean the r^{th} row of the k_1^{th} block matrix row of Ξ , i.e. the r^{th} row of

$$[\Xi_{k_1,1} \quad \dots \quad \Xi_{k_1,d}].$$

Similarly $\text{col}(c, k_2)$ will refer to the c^{th} column of the k_2^{th} block matrix column of Ξ . We will say that the matrix Ξ is in state n , $n \in [N_i]$, if:

1. $\det(\Xi) = \det(\dot{\mathcal{E}}(\mathbf{x}))$.
2. for each $k_1 \in [d]$ and $q \in [N_i]$, the q^{th} row of $(\Xi_{k_1, k_2} - \Gamma)$, where Γ is as defined in (C.12), is

$$\left[\sum_{l=q}^{N_i-n} c_l(k_1, q, k_1) e_l(a_2, \dots, a_{N_i}) \quad \cdots \quad \sum_{l=q}^{N_i-n} c_l(k_1, q, k_1) e_l(a_1, \dots, a_{N_i-1}) \right].$$

3. for each $k_1, k_2 \in [d]$, such that $k_1 \neq k_2$, and each $q \in [N_i]$, the q^{th} row of Ξ_{k_1, k_2} is

$$\left[\sum_{l=q}^{N_i-n} c_l(k_1, q, k_2) e_l(a_2, \dots, a_{N_i}) \quad \cdots \quad \sum_{l=q}^{N_i-n} c_l(k_1, q, k_2) e_l(a_1, \dots, a_{N_i-1}) \right].$$

From (C.7) and (C.8), it is clear that Ξ , as given in (C.10), is in state 1. To verify (C.11), it suffices to show that the state of Ξ can be changed to N_i using reversible elementary row operations. We do so by describing the reversible row operations needed to put Ξ in state $n+1$ starting from a state n , where $1 \leq n < N_i$.

Suppose that $k_1 = k_2$. Then from the definition of Ξ being in state n , it follows that when $N_i - n < q \leq N_i$, the q^{th} row of $\Xi_{k_1 k_2}$ is given by

$$\left[e_{q-1}(a_2, \dots, a_{N_i}) \quad \cdots \quad e_{q-1}(a_1, \dots, a_{N_i-1}) \right],$$

while, for $1 \leq q \leq N_i - n$, it is given by

$$\left[\sum_{l=q-1}^{N_i-n} c_l(k_1, q, k_1) e_l(a_2, \dots, a_{N_i}) \quad \cdots \quad \sum_{l=q-1}^{N_i-n} c_l(k_1, q, k_1) e_l(a_1, \dots, a_{N_i-1}) \right].$$

On the other hand, when $k_1 \neq k_2$, the q^{th} row, $N_i - n < q \leq N_i$, of $\Xi_{k_1 k_2}$ is $\mathbf{0} \in \mathbb{C}^{N_i}$, while for $1 \leq q \leq N_i - n$, it is given by

$$\left[\sum_{l=q}^{N_i-n} c_l(k_1, q, k_1) e_l(a_2, \dots, a_{N_i}) \quad \cdots \quad \sum_{l=q}^{N_i-n} c_l(k_1, q, k_1) e_l(a_1, \dots, a_{N_i-1}) \right].$$

From these observations, it is clear that the effect of subtracting a scalar multiple of $\text{row}(N_i - n + 1, k)$, for some $k \in [d]$, from any other arbitrary row of Ξ , say $\text{row}(r, k_1)$, where $k_1 \in [d]$ and $r \in [N_i]$, is only on the elements whose column index is one of $\text{col}(1, k), \dots, \text{col}(N_i, k)$.

Now fix an arbitrary $k \in [d]$ and consider the following row operations:

1. $\text{row}(N_i - n, k) \leftarrow \text{row}(N_i - n, k) - c_{N_i-n}(k, N_i - n, k) \times \text{row}(N_i - n + 1, k)$. Leaving all other elements unchanged, this operation changes the $(N_i - n)^{th}$ row of the matrix $\Xi_{k,k}$ to

$$\left[e_{N_i-n-1}(a_2, \dots, a_{N_i}) \quad \cdots \quad e_{N_i-n-1}(a_1, \dots, a_{N_i-1}) \right].$$

2. $\text{row}(q, k) \leftarrow \text{row}(q, k) - c_{N_i-n}(k, q, k) \times \text{row}(N_i - n + 1, k)$ for each $1 \leq q < N_i - n$. For every q , leaving other elements unchanged, this operation changes the q^{th} row of $\Xi_{k,k}$ to

$$\left[\sum_{l=q-1}^{N_i-(n+1)} c_l(k, q, k) e_l(a_2, \dots, a_{N_i}) \quad \cdots \quad \sum_{l=q-1}^{N_i-(n+1)} c_l(k, q, k) e_l(a_1, \dots, a_{N_i-1}) \right].$$

3. $\text{row}(N_i - n, k_1) \leftarrow \text{row}(N_i - n, k_1) - c_{N_i-n}(k_1, N_i - n, k) \times \text{row}(N_i - n + 1, k)$ for each $k_1 \in [d] \setminus k$. For every k_1 , leaving other elements unchanged, this operation changes the $(N_i - n)^{th}$ row of the matrix $\Xi_{k_1, k}$ to $\mathbf{0} \in \mathbb{C}^{N_i}$.

4. $\text{row}(q, k_1) \leftarrow \text{row}(q, k_1) - c_{N_i-1}(k_1, q, k) \times \text{row}(N_i - n + 1, k)$ for each $k_1 \in [d] \setminus k$ and each $1 \leq q < N_i - n$. For every k_1, q , leaving other elements unchanged, this operation changes the q^{th} row of $\Xi_{k_1, k}$ to

$$\left[\sum_{l=q}^{N_i-(n+1)} c_l(k_1, q, k) e_l(a_2, \dots, a_{N_i}) \quad \cdots \quad \sum_{l=q}^{N_i-(n+1)} c_l(k_1, q, k) e_l(a_1, \dots, a_{N_i-1}) \right].$$

At present, the matrices $\Xi_{1,k}, \dots, \Xi_{d,k}$ are in the form as required when the Ξ is in state $n + 1$. By repeating the above operations for each $k \in [d]$, the entire Ξ matrix can thus be put in state $n + 1$ as desired. \square

We now recall some standard results from numerical algebraic geometry. Lemmas C.4 and C.5 may be found in [20], Theorems C.6 and C.7 may be found in [24] while Theorem C.8 may be found in [12] or [17].

DEFINITION C.2. *A set $V \subseteq \mathbb{C}^n$ is said to be Zariski closed if there exist $f_1, \dots, f_k \in \mathbb{C}[x_1, \dots, x_n]$, such that*

$$V = \{\mathbf{x} \equiv (x_1, \dots, x_n) \in \mathbb{C}^n : f_1(\mathbf{x}) = \dots = f_k(\mathbf{x}) = 0\}.$$

DEFINITION C.3. *A set $V \subseteq \mathbb{C}^n$ is said to be Zariski open if its compliment is Zariski closed.*

By replacing \mathbb{C} with \mathbb{R} everywhere in the above definitions, Zariski closed and Zariski open sets of \mathbb{R}^n can be equivalently defined.

To distinguish from the open and closed sets of the usual Euclidean topology, the open and closed sets of the Zariski topology with always have the word *Zariski* prefixed to them.

LEMMA C.4. *A Zariski open subset \mathcal{O} of \mathbb{R}^n (or \mathbb{C}^n) is also open in the usual Euclidean topology of \mathbb{R}^n (or \mathbb{C}^n). Further, if \mathcal{O} is non-empty, then it is also dense in the usual topology.*

LEMMA C.5. *Let \mathcal{O} be a non-empty Zariski open subset of \mathbb{C}^n . Then $\mathcal{O} \cap \mathbb{R}^n$ is a non-empty Zariski open subset of \mathbb{R}^n .*

THEOREM C.6. *For a polynomial system $F(\mathbf{x}) \equiv (f_1(\mathbf{x}), \dots, f_n(\mathbf{x})) = 0$, where $f_k : \mathbb{C}^n \rightarrow \mathbb{C}$, the total number of its isolated solutions, counting multiplicities, is bounded above by its total degree, i.e., the product $\deg(f_1) \cdots \deg(f_n)$.*

THEOREM C.7. *For $k \in [n]$, let $f_k(\mathbf{x}; \mathbf{q}) : \mathbb{C}^n \times \mathbb{C}^m \rightarrow \mathbb{C}$ be a polynomial in both \mathbf{x} and \mathbf{q} . Then for the polynomial system $F(\mathbf{x}; \mathbf{q}) = 0$, where $F(\mathbf{x}; \mathbf{q}) \equiv (f_1(\mathbf{x}; \mathbf{q}), \dots, f_n(\mathbf{x}; \mathbf{q}))$, there exists a non-empty Zariski open set $\mathcal{O} \subset \mathbb{C}^m$ such that, for each $\mathbf{q} \in \mathcal{O}$, the system has r_i isolated solutions of multiplicity i , where r_i is an integer independent of $\mathbf{q} \in \mathcal{O}$.*

THEOREM C.8. *Let $F : \mathbb{C}^{d \cdot N_i} \rightarrow \mathbb{C}^{d \cdot N_i}$ be a polynomial map. Then there exists a non-empty Zariski open set U of the co-domain such that for any $\mathbf{u} \in U$, if \mathbf{z} is a root of $F(\mathbf{x}) - \mathbf{u} = 0$, then $\dot{F}(\mathbf{z})$ is non-singular.*

We are now ready to prove Lemma 4.4. But we first show its generalization in $\mathbb{C}^{d \cdot N_i}$.

LEMMA C.9. *There exists an open dense set \mathcal{C}_i of $\mathbb{C}^{d \cdot N_i}$ such that if $\mathbf{w} \in \mathcal{C}_i$, then the solution set of the EPS given in (4.13) satisfies the following properties:*

1. $|V(H_i)| = k \times N_i!$, where $k \in \mathbb{Z}_{++}$ is independent of $\mathbf{w} \in \mathcal{C}_i$, and
2. *Each solution is non-singular.*

Proof. Consider the Zariski open set $\mathcal{C}^1 := \{\mathbf{x} \in \mathbb{C}^{d \cdot N_i} : \det(\dot{\mathcal{E}}(\mathbf{x})) \neq 0\}$. From Lemmas C.1 and C.4, it is clear that \mathcal{C}^1 is non-empty and hence an open dense subset of $\mathbb{C}^{d \cdot N_i}$.

For the map \mathcal{E} that is used to define the EPS, let $U \subseteq \mathbb{C}^{d \cdot N_i}$ be the non-empty Zariski open set as guaranteed in Theorem C.8. Also, let $\mathcal{C}^2 := \mathcal{E}^{-1}(U)$. Clearly, if $\mathbf{w} \in \mathcal{C}^2$, then the EPS necessarily has only non-singular roots. Thus, $\mathcal{C}^2 \subseteq \mathcal{C}^1$. But we now show that \mathcal{C}^2 itself is an open dense subset of $\mathbb{C}^{d \cdot N_i}$. Since U is open and \mathcal{E} is continuous, it follows that \mathcal{C}^2 is open. To show that it is dense we assume the contrary. Then clearly, there exists a $\delta > 0$ and $\mathbf{x} \in \mathbb{C}^{d \cdot N_i} \setminus \mathcal{C}^2$ such that $\mathcal{B}(\mathbf{x}; \delta) \cap \mathcal{C}^2 = \emptyset$. This implies that

$$(C.13) \quad \mathcal{E}(\mathcal{B}(\mathbf{x}; \delta)) \cap U = \emptyset.$$

But \mathcal{C}^1 is an open dense subset. Thus, $\mathcal{B}(\mathbf{x}; \delta) \cap \mathcal{C}^1 \neq \emptyset$. From the inverse function theorem, it follows that, for any $\mathbf{z} \in \mathcal{B}(\mathbf{x}; \delta) \cap \mathcal{C}^1$, \mathcal{E} is a local homeomorphism at \mathbf{z} . Thus, $\mathcal{E}(\mathcal{B}(\mathbf{x}; \delta))$ has non-empty interior. Equation (C.13) then contradicts the fact that U is an open dense subset. Hence, \mathcal{C}^2 is also open dense.

For the map $F(\mathbf{x}, \mathbf{w}) : \mathbb{C}^{d \cdot N_i} \rightarrow \mathbb{C}^{d \cdot N_i}$ defined by $F(\mathbf{x}, \mathbf{w}) = \mathcal{E}(\mathbf{x}) - \mathcal{E}(\mathbf{w})$, let \mathcal{C}^3 be the set as guaranteed in Theorem C.7. Then clearly, $\mathcal{C}_i := \mathcal{C}^2 \cap \mathcal{C}^3$ is an open dense subset of $\mathbb{C}^{d \cdot N_i}$. Furthermore, if $\mathbf{w} \in \mathcal{C}_i$, then

- i. the EPS has only non-singular roots or equivalently $r_i = 0, \forall i \neq 1$, and
- ii. there are precisely $r_1 \in \mathbb{Z}_{++}$ non-singular roots, where r_1 is independent of $\mathbf{w} \in \mathcal{C}_i$.

Note that $r_1 > 1$ since \mathbf{w} is always a root. Because of Theorem C.6, it is also finite.

From the symmetry of the EPS (see Lemma 4.3) and the non-singularity of the roots, it is now easy to see that $|V(H_i)| = k \times N_i!$ for some $k \in \mathbb{Z}_{++}$. \square

Let $\mathcal{C}^1, \mathcal{C}^2, \mathcal{C}^3$ and U be defined as in the proof above. From Lemmas C.4 and C.5, it follows that $\mathcal{R}^1 := \mathcal{C}^1 \cap \mathbb{R}^{d \cdot N_i}$, $U \cap \mathbb{R}^{d \cdot N_i}$ and $\mathcal{R}^3 := \mathcal{C}^3 \cap \mathbb{R}^{d \cdot N_i}$ are open dense subsets of $\mathbb{R}^{d \cdot N_i}$. The same argument that was used to prove \mathcal{C}^2 is an open dense subset of $\mathbb{C}^{d \cdot N_i}$, with the real version of the inversion of the inverse function theorem, also shows that $\mathcal{R}^2 := \mathcal{E}^{-1}(U \cap \mathbb{R}^{d \cdot N_i}) \cap \mathbb{R}^{d \cdot N_i}$ is an open dense subset of $\mathbb{R}^{d \cdot N_i}$. But note that all coefficients of \mathcal{E} are real. Hence, $\mathcal{R}^2 = \mathcal{E}^{-1}(U) \cap \mathbb{R}^{d \cdot N_i} = \mathcal{C}^2 \cap \mathbb{R}^{d \cdot N_i}$. This now implies that $\mathcal{R}_i := \mathcal{R}^2 \cap \mathcal{R}^3 = \mathcal{C}_i \cap \mathbb{R}^{d \cdot N_i}$ is an open dense subset of $\mathbb{R}^{d \cdot N_i}$. Since $\mathcal{R}_i \subset \mathcal{C}_i$, even if $\mathbf{w} \in \mathcal{R}_i$ the EPS has the same properties as those described in Lemma C.9. This now completes the verification Lemma 4.4 as desired.

Acknowledgments. The author would like to sincerely thank his advisors, Prof. V. Borkar and Prof. D. Manjunath, for inspiring, motivating and guiding him right through this work. He would also like to thank C. Wampler, A. Sommesse, R. Gandhi, S. Gurjar and M. Gopalkrishnan for helping him understand several results from algebraic geometry. Lastly, he would like to thank S. Juneja for referring [2].

REFERENCES

- [1] A. ADAMS, T. BU, R. CACERES, N. DUFFIELD, T. FRIEDMAN, J. HOROWITZ, F. LO PRESTI, S. B. MOON, V. PAXSON, AND D. TOWSLEY, *The use of end-to-end multicast measurements for characterizing internal network behaviors*, IEEE Communications Magazine, (2000).
- [2] R. F. BOTHA AND C. M. HARRIS, *Approximation with generalized hyperexponential distributions: Weak convergence results*, Queueing Systems 2, (1986), pp. 169 – 190.
- [3] R. CACERES, N. G. DUFFIELD, J. HOROWITZ, AND D. F. TOWSLEY, *Multicast-based inference of network-internal loss characteristics*, IEEE Transactions on Information Theory, 45 (1999), pp. 2462–2480.
- [4] R. CASTRO, M. J. COATES, G. LIANG, R. NOWAK, AND B. YU, *Internet tomography: Recent developments*, Statistical Science, 19 (2004), pp. 499–517.
- [5] A. CHEN, J. CAO, AND T. BU, *Network tomography: Identifiability and fourier domain estimation*, IEEE Transactions on Signal Processing, 58 (2010), pp. 6029–6039.

- [6] Y. CHEN, D. BENDEL, AND R.H. KATZ, *An algebraic approach to practical and scalable overlay network monitoring*, in Proceedings of IEEE INFOCOM, 2001.
- [7] M. J. COATES AND R. NOWAK, *Network tomography for internal delay estimation*, in Proceedings of IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP), 2001.
- [8] K. DENG, Y. LI, W. ZHU, Z. GENG, AND J.S. LIU, *On delay tomography: fast algorithms and spatially dependent models*, IEEE Transactions on Signal Processing, 60 (2012), pp. 5685 – 5697.
- [9] B. K. DEY, D. MANJUNATH, AND S. CHAKRABORTY, *Estimating network link characteristics using packet-pair dispersion: A discrete-time queueing theoretic analysis*, Computer Networks, 55 (2011), pp. 1052–1068.
- [10] M. H. FIROOZ AND S. ROY, *Network tomography via compressive sensing*, Proc. of IEEE Globecom, (2010).
- [11] O. GUREWITZ AND M. SIDI, *Estimating one-way delay from cyclic-path delay measurements*, in Proceedings of IEEE INFOCOM, 2001.
- [12] R. HARTSHORNE, *Algebraic Geometry*, Springer-Verlag, 1977.
- [13] W. Hoeffding, *Probability inequalities for sums of bounded random variables*, J. Amer. Statist. Assoc., (1963).
- [14] E. LAWRENCE, G. MICHAELIDIS, AND V. N. NAIR, *Statistical inverse problems in active network tomography*, Lecture-Notes Monograph Series, 54, Complex Datasets and Inverse Problems: Tomography, Networks and Beyond (2007).
- [15] L. J. LEBLANC AND K. FARHANGIAN, *Selection of trip matrices from observed link volumes*, Transportation Research, 16B (1982).
- [16] T. Y. LI, *Numerical solution of multivariate polynomial systems by homotopy continuation methods*, Acta Numerica, (1997), p. 399.
- [17] T. Y. LI AND X. WANG, *The BKK root count in \mathbb{C}^n* , Math. Comp, 65 (1996), pp. 1477–1484.
- [18] X. LIU, K. RAVINDRAN, AND D. LOGUNOV, *A queueing-theoretic foundation of available bandwidth estimation: Single-hop analysis*, IEEE/ACM Transactions on Networking, 15 (2007), pp. 918–931.
- [19] A. MORGAN, A. SOMMESE, AND L. WATSON, *Finding all isolated solutions to polynomial systems using hompack*, ACM Trans. Math. Softw., 15 (1989), pp. 93 – 122.
- [20] D. MUMFORD, *The red book of varieties and schemes*, Springer-Verlag, 1999, pp. 57–63.
- [21] J. OU, J. LI, AND S. OZEKICI, *Approximating a cumulative distribution function by generalized hyperexponential distributions*, Probability in the Engineering and Informational Sciences, (1997), pp. 11 – 18.
- [22] F. LO PRESTI, N. G. DUFFIELD, J. HOROWITZ, AND D. TOWSLEY, *Multicast-based inference of network-internal delay distributions*, IEEE/ACM Transactions on Networking, 10 (2002), pp. 761–775.
- [23] M. F. SHIH AND A. O. HERO, *Unicast-based inference of network link delay distributions using mixed finite mixture models*, IEEE Transactions on Signal Processing, 51 (2003), pp. 2219–2228.
- [24] A. J. SOMMESE AND C. W. WAMPLER, *The numerical solution of systems of polynomials*, World Scientific, 2005.
- [25] G. THOPPE, *Generalized network tomography*, in Proceedings of 50th Annual Allerton Conference on Communication, Control and Computing, 2012.
- [26] Y. TSANG, M. J. COATES, AND R. NOWAK, *Network delay tomography*, IEEE Transactions on Signal Processing: Special Issue on Signal Processing in Networking, 51 (2003).
- [27] Y. VARDI, *Network tomography: Estimating source-destination traffic intensities from link data*, Journal of the American Statistical Association, 91 (1996), pp. 365–377.
- [28] J. VERSCHELDE, *Algorithm 795: Phcpack: a general-purpose solver for polynomial systems by homotopy continuation*, ACM Trans. Math. Softw., 25 (1999), pp. 251–276.
- [29] Y. ZHANG, M. ROUGHAN, C. LUND, AND D. DONOHO, *An information-theoretic approach to traffic matrix estimation*, in Proceedings of ACM SIGCOMM, August 25–29 2003.